



Inférence statistique par des transformées de Fourier pour des modèles de régression semi-paramétriques

Myriam Vimond

► To cite this version:

Myriam Vimond. Inférence statistique par des transformées de Fourier pour des modèles de régression semi-paramétriques. Mathématiques [math]. Université Paul Sabatier - Toulouse III, 2007. Français. NNT: . tel-00185102

HAL Id: tel-00185102

<https://theses.hal.science/tel-00185102>

Submitted on 5 Nov 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Laboratoire de Statistique et Probabilités

Inférence statistique par des transformées de Fourier pour des modèles de régression semi-paramétriques

THÈSE

présentée et soutenue publiquement le 12 juillet 2007

pour l'obtention du

Doctorat de l'université Paul Sabatier – Toulouse III

(spécialité probabilités et statistique)

par

Myriam Vimond

Composition du Jury

<i>Directeur de thèse :</i>	Fabrice Gamboa	Toulouse 3
<i>Rapporteurs :</i>	Marc Hallin	Université libre de Bruxelles
	Sarah van der Geer	ETH Zürich
<i>Examineur :</i>	Pascal Massart	Paris XI
	Béatrice Laurent	Toulouse 3
	Jean-Michel Loubes	Montpellier 2

Inférence statistique par des transformées de Fourier pour des modèles de régression semi-paramétriques

par

Myriam Vimond



*Ne vous mettez pas à la torture pour le lendemain.
Demain pensera à lui-même, et vous y penserez quand vous le vivrez.
Laissez à chaque jour ses ennuis !*

Maria Valtorta, tome 3, chap. 33

Table des matières

Table des matières	i
1 Outils d'analyse statistique et d'analyse de Fourier	11
1.1 M -estimation	12
1.2 Efficacité dans des modèles semi-paramétriques	14
1.2.1 Estimation semi-paramétrique efficace	15
1.2.2 Vraisemblance profilée	18
1.3 Méthode bootstrap	19
1.4 Transformée de Fourier	21
2 Estimation pour des modèles de forme invariante	25
2.1 Présentation du modèle	28
2.1.1 Description du modèle	29
2.1.2 Identifiabilité du modèle	31
2.2 Définition des processus de contraste	32
2.2.1 Transformée de Fourier discrète du modèle	32
2.2.2 Définition empirique d'un processus de contraste.	34
2.2.3 Définition d'un processus de contraste à partir de la vraisem- blance profilée	38
2.3 Efficacité et normalité asymptotique	43
2.3.1 Normalité Asymptotique Locale (LAN) du modèle	43
2.3.2 Estimation efficace du paramètre α^*	45
2.3.3 Estimation de la forme commune f^*	51
2.4 Application des méthodes d'estimation	53
2.4.1 Les contraintes d'identifiabilité	53
2.4.2 Choix de la suite de poids.	55
2.4.3 Choix du nombre de coefficients de Fourier m_n	57

2.5	Preuves de la Section 2.2	58
2.6	Preuves de la Section 2.3	70
2.6.1	Preuve du théorème 2.4	74
2.6.2	Preuves et lemmes techniques concernant l'estimateur de la forme commune	78
3	Test d'adéquation pour un modèle de forme invariante	81
3.1	Estimation des paramètres	84
3.1.1	Description du modèle (3.2)	84
3.1.2	Estimation des paramètre pour le modèle (3.2).	85
3.1.3	Estimation des paramètres pour le modèle général (3.1)	89
3.1.4	Application de la méthode d'estimation sur les ondes cérébrales P300	91
3.2	Test d'adéquation du modèle	92
3.2.1	Les statistiques de test.	93
3.2.2	La méthode bootstrap	95
3.2.3	Les simulations	96
3.2.4	Preuves	100
3.3	Test sur l'homogénéité des données	102
3.3.1	Estimation des paramètres de translation	103
3.3.2	Test statistique sur un coefficient de Fourier	106
3.3.3	Agrégation des statistiques de test	108
3.3.4	Preuves	110
4	Estimation des $\mathcal{R.S.T.}$ entre des images bruitées	117
4.1	Un modèle semi-paramétrique pour l'alignement affine	120
4.1.1	Description du modèle	121
4.1.2	Identifiabilité du modèle	124
4.2	Transformées de Fourier et invariance par $\mathcal{R.S.T.}$	125
4.2.1	Transformée de Fourier standard	125
4.2.2	Transformée de Fourier-Mellin	125
4.3	Estimation des paramètres	127
4.3.1	Un critère pour l'estimation simultanée des échelles et des angles de rotation	129
4.3.2	Un critère pour l'estimation simultanée des vecteurs de translation	135
4.4	Simulations	137
4.4.1	Un exemple réel	140
4.5	Preuves	140

5	Alignement de séries temporelles	157
5.1	Présentation du modèle	160
5.2	Estimation des décalages	161
5.2.1	Estimation par processus de contraste	161
5.2.2	Simulations	165
5.3	Propositions et lemmes techniques	171
	Bibliographie	179
A	Notions et notations utilisées	187
A.1	Rappels de probabilité et statistiques	187
A.1.1	Variables gaussiennes complexes	187
A.1.2	Convergence Stochastique	187
A.2	Espace de Hilbert	188
A.3	Produit tensoriel de deux espaces de dimensions finies	189
B	Programmes MATLAB	191
B.1	Programmes relatifs au chapitre 2	191
B.1.1	Calcul des estimateurs	191
B.1.2	Définition du processus de contraste (vraisemblance profilée) . .	192
B.2	Programmes relatifs au chapitre 3	193
B.2.1	Algorithme bootstrap du test	193
B.2.2	Définition du processus de contraste	195
B.3	Programmes relatifs au chapitre 5	195
B.3.1	Définition du processus de contraste	196
	Liste des symboles et abréviations	197
	Table des figures	199
	Liste des tableaux	201

Remerciements

Je voudrais tout d'abord remercier Fabrice GAMBOA qui, en me proposant de commencer une thèse, m'a permise de goûter aux joies de la recherche et de l'enseignement. Je le remercie pour sa confiance, son soutien et ses conseils, pour la direction de ce travail de recherche.

Je suis très honorée par la présence de Marc HALLIN dans mon jury. Je le remercie ainsi que Sarah VAN DER GEER d'avoir accepté d'être rapporteurs de cette thèse.

Je suis très reconnaissante à Pascal MASSART d'être dans mon jury de thèse. Sans ses encouragements, durant l'année de DEA à Orsay, je n'aurais sans doute pas persévéré dans mes études.

Je remercie Béatrice LAURENT d'avoir si volontiers accepté de faire partie de mon jury de thèse et pour ses questions qui m'ont permise de pousser un peu plus loin mes résultats.

Je tiens à remercier Jean-Michel LOUBES pour sa présence dans le jury, d'autant plus qu'il est l'un des initiateurs de ce sujet de thèse.

Un grand merci à Jérémie BIGOT pour sa disponibilité, et pour sa contribution qu'il a apporté à ce travail. Je remercie également Hölger DETTE pour son aide dans mes recherches sur les tests bootstrap, et pour son chaleureux accueil à l'Université de Bochum.

Pendant ces trois années, j'ai effectué mon travail de recherche au sein de l'équipe du LSP (Laboratoire de Statistique et de Probabilités), et j'ai bénéficié des moyens mis à ma disposition par celle-ci. J'ai eu beaucoup de plaisir à participer aux séminaires de

Statistique, au groupe de travail de Statistique Asymptotique et à côtoyer les membres du groupe MAFIA (Modélisation Aléatoire Finalité Industrielle Appliquée). Naturellement, je n'oublie pas tous ceux avec qui j'ai partagé bureau, problèmes pratiques, discussions de travail et pauses chocolat, en particulier : Matthieu et Jean-Paul pour cette dernière année au bureau, Jérôme pour m'avoir "convertie" aux logiciels libres, Delphine, Solenn et Diana pour les deux "saisons" de badminton, Lionel, Christophe, Renaud, Florent, Maxime, Laurent, Amélie, Aurélien, Nolwen, Michel, Erwan, Ignacio, Matthieu, Amandine et Sebastien pour la formidable ambiance et leurs entrains, Agnès pour l'organisation des repas entre doctorants. Un merci particulier à Lionel et à Delphine pour l'aide orthographique

Je n'oublie pas que je suis redevable à tous mes enseignants qui m'ont transmis leurs connaissances et m'ont donné de bonnes leçons de vie.

Mes derniers remerciements vont évidemment aux personnes à qui je tiens le plus et qui m'ont soutenu tout au long de mes travaux de thèse. J'aimerais remercier Claude, Mariama, et toutes les personnes qui pourront se reconnaître dans Loubéiac pour leurs amitiés et leurs réconforts. Je n'exagère pas quand je dis que je laisse, avec vous tous, une part de mon coeur à Toulouse. Enfin, je remercie l'autre partie de mon coeur : ma famille, spécialement mes parents et mon frère pour leurs soutiens constants.

Introduction

Le travail présenté ici s'inscrit dans le cadre de l'étude de modèles semi-paramétriques. Nous étudions des modèles de régression décrivant l'observation simultanée d'un nombre fixé J de signaux discrets ou continus,

$$Y_{i,j} = f_j^*(t_{i,j}) + \sigma_j^* \epsilon_{i,j}, \quad i = 1, \dots, n_j, \quad j = 1 \dots J.$$

Les objets d'intérêt sont les signaux inconnus f_j observés aux points d'observation $t_{i,j} \in \mathcal{D}$ (avec $\mathcal{D} \subset \mathbb{R}^d$ et $d \geq 1$ ou $\mathcal{D} \subset \mathbb{Z}$). Les erreurs $\epsilon_{i,j}$ dues aux mesures sont modélisées par un bruit blanc. Plus particulièrement, on s'est intéressé aux modèles où les signaux f_j ont une forme similaire, c'est-à-dire où ils sont identiques à une déformation près,

$$f_j^* = D_j f^*, \quad j = 1, \dots, J,$$

avec f^* modélise la forme commune du signal, et D_j sont les opérateurs de déformation associés à chacun des signaux observés. De tels modèles trouvent des applications en médecine (Kneip et Gasser (1988), Wang et Brown (1996) ou Laciari et al. (2003)), en traitement de l'image (Glasbey (1998)) ou encore en traitement de données relatives au trafic routier (Loubes et al. (2006)).

Pour ces types de modèles, avant d'exploiter les données, un traitement indispensable est bien souvent l'alignement des signaux observés. L'un des premiers problèmes à traiter est alors le choix du signal de référence sur lequel vont être alignés les signaux f_j^* . Une première idée, dans certains cas, serait de considérer la moyenne des signaux, mais il est bien connu qu'un tel estimateur sous-estime les pentes et les pics de la forme commune (Ramsay et Li (1998)). Il est donc nécessaire au préalable d'estimer les déformations D_j . Dans cette thèse, nous supposons que les opérateurs D_j appartiennent à une famille paramétrique. Les modèles considérés sont donc bien semi-paramétriques : ils sont décrits par un paramètre fini-dimensionnel qui caractérise les déformations, et par un paramètre de nuisance déterminant la forme commune f des

signaux. Ces modèles ont été pour la première fois définis par [Lawton et al. \(1972\)](#) sous le nom de modèles de forme invariante (ou *shape invariant models*).

Pour les différents modèles étudiés dans cette thèse, nous proposons des M -estimateurs pour les paramètres caractérisant les déformations D_j . Ces estimateurs maximisent ou minimisent des fonctions dites de contraste construites à partir des transformées de Fourier des données. La transformée de Fourier possède, en effet, la propriété remarquable de diagonaliser simultanément tous les opérateurs de translation. Le choix du groupe, sur lequel la transformée de Fourier sera considérée, dépendra de la nature de la famille des transformations étudiées. Nos méthodes ne nécessitent pas de signal de référence pour aligner les J signaux : nous considérons la moyenne des signaux synchronisés et fixons des contraintes d'identifiabilité. Nous étudions également dans le cas $d = 1$ un test d'adéquation construit à partir d'une de ces fonctions de contraste.

Le plan de cette thèse est le suivant. Dans le chapitre 1, nous rappelons principalement quelques éléments sur la théorie de l'efficacité dans des modèles semi-paramétriques, et sur la transformée de Fourier. Les chapitres 2 et 4 étudient le problème d'alignement de signaux continus pour deux familles de déformations différentes. Par ailleurs, nous montrons au chapitre 2 que les estimateurs proposés sont asymptotiquement efficaces. Dans le chapitre 3, on construit des procédures de test d'adéquation pour les modèles étudiés au chapitre précédent. Enfin, dans le chapitre 5, nous étudions le problème d'alignement de séries temporelles ($\mathcal{D} \subset \mathbb{Z}$). Tous les estimateurs proposés sont consistants et leurs vitesses de convergence sont étudiées.

Nous donnons un bref aperçu de nos résultats.

Chapitre 2

La première famille d'opérateurs de déformation sur des signaux continus $f : \mathbb{R}^d \rightarrow \mathbb{R}$ ($d \geq 1$) est la suivante,

$$D : f \rightarrow Df(\cdot) = af(\cdot - \theta) + v,$$

où $a \in \mathbb{R}^*$ modifie l'amplitude, $\theta \in \mathbb{R}^d$ est un paramètre de translation et $v \in \mathbb{R}$ modifie le niveau. Ce type de modèle a été proposé et étudié pour la première fois pour $d = 1$ par [Lawton et al. \(1972\)](#). Dans ce dernier article, les auteurs développent un algorithme itératif, SEMOR, pour estimer simultanément les paramètres (θ, a, v) et la forme commune f^* . La consistance de cette méthode a ensuite été prouvée par [Kneip et Gasser \(1988\)](#). [Härdle et Marron \(1990\)](#) définissent des estimateurs (de θ et v) asymptotiquement normaux de vitesse \sqrt{n} construits à partir d'un estimateur à noyau de la forme commune. Plus récemment, toujours pour $d = 1$, [Gamboa et al. \(2005\)](#)

construisent un estimateur asymptotiquement normal de vitesse \sqrt{n} du paramètre de translation θ à partir de la transformée de Fourier discrète (DFT) des données.

Nos résultats

Les méthodes d'estimation des paramètres (θ, a, v) proposées ici généralisent pour une dimension quelconque d celle proposée par [Gamboa et al. \(2005\)](#). Les estimateurs sont tous asymptotiquement normaux de vitesse \sqrt{n} (théorème 2.2), voire asymptotiquement efficaces (théorème 2.4) pour certains d'entre eux. Nous définissons exactement trois estimateurs pour (θ, a, v) associés respectivement à trois jeux de contraintes d'identifiabilité différentes (section 2.1.2) et donc à trois critères différents (voir les équations (2.16), (2.26) et (2.30)). Ces critères sont définis à partir de la DFT des données. En effet la transformée de Fourier sur le tore est, par sa linéarité et par sa propriété de diagonalisation sur les opérateurs de translation, bien adaptée à la famille de déformations étudiée (section 2.2.1).

Deux familles de critères sont considérées. La première utilise une suite de poids (2.16) qui pondère les coefficients de Fourier. Elle se situe dans la lignée des travaux de [Dalalyan et al. \(2006\)](#). La seconde famille, (2.26) et (2.30), est définie à partir de la vraisemblance profilée. L'étude de cette méthode conduit à des calculs de coefficients de Fourier empiriques comme ceux considérés dans [Gassiat et Lévy-Leduc \(2006\)](#). Suite aux résultats établis, on peut faire les remarques suivantes :

- l'emploi d'une suite de poids régulière permet d'exiger très peu de régularité sur la forme commune, toute la régularité étant reportée sur le poids ;
- la vraisemblance profilée impose des contraintes d'identifiabilité sur le modèle (remarque 2.5) et conduit comme l'ont établi [Murphy et Van der Vaart \(2000\)](#) (pour des modèles semi-paramétriques pour des observations i.i.d.) à estimer efficacement les paramètres (théorème 2.4) ;
- les hypothèses des théorèmes 2.3 et 2.4 signifient que plus la dimension d de l'espace d'observation augmente, plus on doit considérer des fonctions f_j régulières (une différentiabilité supérieure à $d/2$ au moins). Ceci a des conséquences lorsque l'on désire établir les vitesses de convergence de la forme commune (corollaire 2.1) ;

Perspectives

Les choix de la suite de poids ou du nombre de coefficients de Fourier ne sont pas entièrement déterminés par les hypothèses des différents résultats de convergence. Le chapitre 2 se termine ainsi par une étude basée sur des simulations concernant ce

choix. En effet, si les estimateurs de (θ, a, v) construits sont tous asymptotiquement centrés, il existe un biais pour n fixé. La suite de poids ou le nombre de coefficients de Fourier considérés doit donc réaliser un compromis entre le biais, la régularité de la forme commune et le rapport signal sur bruit. Cette remarque conduit à étudier le problème de l'estimation des paramètres sous l'angle de la sélection de modèle. Ce travail est en cours de réalisation en collaboration avec Béatrice LAURENT et Matthieu LERASLE

Chapitre 3

Parallèlement à l'étude des estimations des paramètres dans des modèles de forme invariante, les statisticiens se sont intéressés aux problèmes d'adéquation du modèle aux données. Möcks et al. (1984) proposent deux procédures de test visant à tester si les fonctions f_j sont identiques. Pour des sous modèles de celui considéré dans ce chapitre, Härdle et Marron (1990) construisent des procédures afin de tester l'adéquation aux données et l'élimination d'un des paramètres. Leurs statistiques de test sont définies à partir de leurs critères d'estimation.

Nos résultats

Nous proposons deux procédures de test. La première teste exactement l'adéquation des modèles étudiés au chapitre 2 pour $d = 1$. La seconde teste une hypothèse d'homogénéité sur les coefficients de Fourier des fonctions f_j . Les statistiques de test sont dans les deux cas définies à partir de critères similaires à ceux utilisés pour l'estimation (section 3.1).

La première procédure teste l'hypothèse suivante :

$$H_0 : \quad \text{il existe une fonction } f \text{ telle que pour tout } j = 1, \dots, J, \text{ il existe } a_j \in \mathbb{R}^*, \\ \theta_j \in \mathbb{R} \text{ et } v_j \in \mathbb{R} \text{ avec } f_j^*(\cdot) = a_j f(\cdot - \theta_j) + v_j,$$

contre,

$$H_1 : \quad \text{pour toute fonction } f, \text{ il existe } j = 1, \dots, J, \text{ tel que pour tout } a_j \in \mathbb{R}^*, \theta_j \in \mathbb{R} \\ \text{et } v_j \in \mathbb{R} \text{ il existe } t \in \mathbb{R} \text{ et } f_j^*(t) \neq a_j f(t - \theta_j) + v_j.$$

Du fait de l'utilisation des coefficients de Fourier, on est amené à supposer la périodicité des fonctions f_j . Ce test est donc parfaitement adapté pour tester l'existence de phénomènes cycliques, comme les rythmes biologiques (Wang et Brown (1996)). Il peut être aussi appliqué à des cas plus généraux (voir remarque 3.1). La statistique

de test (corollaire 3.1) est définie comme la borne inférieure d'un critère d'estimation : quand l'hypothèse nulle est vraie, cette statistique converge en loi. Dans le cas contraire elle tend vers l'infini avec n . Puisque la loi limite de la statistique de test est compliquée sous l'hypothèse nulle, le test est implémenté par une procédure bootstrap. Des simulations, illustrant les performances de la procédure bootstrap, montrent que le test est assez sensible : il suffit d'une légère modification dans la forme ou d'un petit nombre de signaux ne correspondant pas au modèle pour rejeter l'hypothèse nulle.

La seconde procédure teste l'homogénéité des fonctions de régression au sens suivant : l'hypothèse nulle est,

$$H_0 : \quad \text{pour tout } l \in \mathbb{Z}, \text{ et pour tout } (j, k) \in \{1 \dots J\}^2, |c_l(f_j^*)| = |c_l(f_k^*)|,$$

contre,

$$H_1 : \quad \text{il existe } l \in \mathbb{Z}, \text{ et } (j, k) \in \{1 \dots J\}^2 \text{ tels que } |c_l(f_j^*)| \neq |c_l(f_k^*)|.$$

Ici $c_l(f)$ désigne le coefficient de Fourier d'ordre l de la fonction f . Cette dernière hypothèse nulle contient le cas où les fonctions f_j^* diffèrent seulement les unes des autres d'un facteur de translation. La statistique de test (théorème 3.3) est le maximum d'un ensemble croissant de statistiques indépendantes. Chacune de ces statistiques est définie à partir d'un critère d'estimation (similaires à ceux définis dans le chapitre 2) ne prenant en compte qu'un seul coefficient de Fourier à la fois. Sous l'hypothèse nulle, ces statistiques sont identiquement distribuées (théorème 3.2). Par conséquent la convergence de la statistique de test se déduit de la théorie des valeurs extrêmes (voir Resnick (1987)). Pour utiliser cette procédure, la taille de l'échantillon doit donc être assez grande. Nous obtenons néanmoins des simulations assez convaincantes pour des tailles d'échantillons raisonnables.

Perspectives

La procédure de test bootstrap est étudiée dans ce chapitre uniquement d'un point de vue numérique. La loi limite sous l'hypothèse nulle dépend des paramètres (des coefficients de Fourier de la forme commune notamment) de façon assez complexe. Les travaux de Beran (1988) et de Fisher et Hall (1990) suggèrent que la statistique de test devrait être asymptotiquement un pivot pour bien fonctionner. Une étude statistique (plus approfondie) de la puissance du test proposé sera menée plus tard.

Chapitre 4

Le chapitre 4 est un travail en collaboration avec Jérémie BIGOT et Fabrice GAMBOA. On étudie le problème d'alignement d'images bidimensionnelles. On considère ici la famille de déformations suivante sur les signaux continus $f : \mathbb{R}^2 \rightarrow \mathbb{R}$,

$$D : f \rightarrow Df(\cdot) = f\left(\frac{1}{a}R_\theta(\cdot - b)\right),$$

où $a \in \mathbb{R}_+^*$ modifie l'échelle de l'image, R_θ est la matrice de rotation d'angle $\theta \in [0, 2\pi]$ et $b \in \mathbb{R}^2$ est un vecteur de translation. Ce groupe de déformations est l'ensemble des similitudes directes : elles conservent l'orientation et le rapport des distances. Pour comparer deux images différant de la sorte, la méthode la plus couramment utilisée est la technique de corrélation de phase (Kuglin et Hines (1975)) basée sur la propriété des transformées de Fourier des fonctions translatées. D'autre part en récrivant les fonctions en coordonnées log-polaires, l'opérateur de rotation et d'homothétie se traduisent en un opérateur de translation pour ce système de coordonnées. C'est pourquoi, beaucoup de méthodes d'estimation des paramètres des similitudes sont basées sur ce principe et procèdent généralement en deux étapes. La première étape utilise le module de la transformée de Fourier des images exprimée en coordonnée log-polaire pour estimer le rapport d'homothétie et l'angle de rotation. La seconde étape estime le vecteur de translation en réalignant les transformées de Fourier des images (voir par exemple Reddy et Chatterji (1996), Keller et al. (2005a)).

Nos résultats

Nous développons un procédé d'estimation en deux étapes pour aligner un ensemble fini d'images bruitées qui sont identiques à une similitude près et dont la forme commune est inconnue. Les deux critères proposés sont construits à partir de la transformée de Fourier et de la transformée de Fourier-Mellin (Derrode et Ghorbel (2004)). Ils fournissent des estimateurs consistants (théorèmes 4.1 et 4.3) dont nous avons étudié la vitesse de convergence (théorèmes 4.2 et 4.4). Nous nous sommes placés ici dans un modèle semi-paramétrique de bruit blanc et non plus dans un modèle de régression comme pour les chapitres précédents. La modélisation de type bruit blanc fournit, en effet, un cadre théorique pour l'étude de problèmes de régression d'images 2D (cf. Brown et Low (1996) et Donoho et Johnstone (1999)). Il faut souligner que les problèmes d'interpolation des images compliquent significativement l'étude sur la consistance des estimateurs que nous proposons ici.

Perspectives

En comparant avec l'étude faite au chapitre 2, la question de l'efficacité des estimateurs se pose et plus particulièrement pour l'estimateur des paramètres d'échelle et de rotation. En effet lorsque l'on considère le module de la transformée de Fourier en coordonnée log-polaire, la famille de déformation étudiée n'est pas exactement celle étudiée au chapitre 2 :

$$D : f \rightarrow Df(\cdot, \cdot) = a^4 f(\cdot + \log(a), \cdot + \theta), \quad a \in \mathbb{R}_+^*, \theta \in [0, 2\pi].$$

Ici le paramètre de translation est fonction du paramètre d'amplitude. L'étude de la borne de Cramer Rao asymptotique est un axe de recherche que nous comptons développer par la suite.

Chapitre 5

Le dernier chapitre est consacré aux opérateurs de translation pour des signaux discrets ($\mathcal{D} \subset \mathbb{Z}$),

$$D : f = (f_l)_{l \in \mathbb{Z}} \rightarrow Df = (f_{l-\theta})_{l \in \mathbb{Z}},$$

où $\theta \in \mathbb{Z}$ est un paramètre de décalage. Pour aligner deux signaux de ce type, il existe un large éventail de méthodes comme la corrélation croisée (Laciar et al. (2003)), des techniques utilisant des intégrales normalisées (Jane et al. (1991)) ou des chemins dynamiques (Sakoe et Chiba (1978)).

Nos résultats

L'étude statistique de ce chapitre est duale à celle menée dans le chapitre 2. En effet, la transformée de Fourier établit une dualité entre les fonctions périodiques de carré intégrable et les séries de carré sommable. Dans le chapitre 2, les paramètres de translations sont estimés à partir des coefficients de Fourier des signaux continus. Ici, les paramètres de décalage sont estimés à partir des fonctions 2π périodiques associées aux signaux discrets observés. Le critère d'estimation des décalages est donc fonction des autopériodogrammes (le module au carré des transformées de Fourier). On s'est intéressé aux signaux ergodiques, c'est-à-dire aux signaux admettant une mesure spectrale. Le critère proposé fournit des estimateurs consistants (théorème 5.1). Il est original dans le sens où il n'est pas seulement défini sur les entiers relatifs mais sur les réels. Lorsque l'optimisation du critère est restreinte aux entiers relatifs, la méthode d'estimation généralise pour un nombre fixé de signaux la technique de corrélation croisée (Laciar et al. (2003)).

Perspectives

Suite aux résultats établis et aux simulations, deux pistes se dégagent en vue d'améliorer le critère. Premièrement, pour assurer la consistance de l'estimateur, nous avons supposé que la variance du bruit converge en moyenne de Cesaro vers zéro (hypothèse (5.2)). Cette hypothèse limite la robustesse du critère, et impose une vitesse de convergence inférieure à \sqrt{n} (théorème 5.2). Deuxièmement, le fait de pouvoir définir le critère sur les réels et non uniquement sur les entiers relatifs pouvait faire espérer l'emploi d'outil d'optimisation différentielle pour le calcul des estimateurs. Or si le critère limite possède bien un unique maximum global au point désiré, il possède aussi une multiplicité de maximum locaux. Ce problème peut être évité par l'insertion d'une fonction de poids qui relocalise le critère. Mais il reste encore à déterminer quelle fonction de poids choisir, et quelle est son influence sur la qualité de l'estimation.

Chapitre 1

Outils d'analyse statistique et d'analyse de Fourier

Sommaire

1.1	<i>M</i> -estimation	12
1.2	Efficacité dans des modèles semi-paramétriques	14
1.3	Méthode bootstrap	19
1.4	Transformée de Fourier	21

Introduction

Dans ce chapitre, les principaux outils d'analyse statistique et d'analyse de Fourier utilisés dans cette thèse sont introduits. Tous les modèles que nous allons étudier sont des modèles semi-paramétriques. Soit Θ un ensemble de paramètres quelconque. Un modèle statistique \mathcal{P} est une collection de mesures de probabilités sur un espace mesurable $(\mathcal{X}, \mathcal{T})$,

$$\mathcal{P} = \{\mathbb{P}_\vartheta, \vartheta \in \Theta\}.$$

Supposons que l'on observe k_n variables aléatoires dans ce modèle :

$$X_{1,n}, \dots, X_{k_n,n} \sim \mathbb{P}_\vartheta^{(n)},$$

où $\mathbb{P}_\vartheta^{(n)} = \mathbb{P}_\vartheta|_{\mathcal{T}_n}$ est la restriction de \mathbb{P}_ϑ à la tribu d'évènements \mathcal{T}_n engendrée par les variables $X_{1,n}, \dots, X_{k_n,n}$. On note par $\mathcal{P}^{(n)}$ l'ensemble des mesures de probabilité suivant :

$$\mathcal{P}^{(n)} = \{\mathbb{P}_\vartheta^{(n)}, \vartheta \in \Theta\}.$$

Ce modèle est dit semi-paramétrique lorsque le paramètre ϑ peut s'écrire sous la forme $\vartheta = (\varphi, \eta)$, où $\varphi \in \Theta_\varphi \subset \mathbb{R}^d$ est la composante fini-dimensionnelle, et $\eta \in \Theta_\eta$ est la composante infini-dimensionnelle, appelée aussi nuisance. Dans un tel modèle, le paramètre d'intérêt est souvent le paramètre fini-dimensionnel φ , que l'on cherche à estimer.

Les estimateurs que nous étudions sont des M -estimateurs, c'est-à-dire qu'ils maximisent ou minimisent une fonction définie à partir des données. La théorie et la terminologie employées sont rappelées dans la section 1.1. Dans cette thèse, ces fonctions sont définies à partir d'une transformée de Fourier des données. Suivant la géométrie du modèle, on utilise des transformées de Fourier différentes. La section 1.4 rappelle les grandes lignes de la théorie de Fourier sur les groupes abéliens localement compact. Cette théorie permet de donner un cadre unifié pour les différentes transformées de Fourier utilisées dans cette thèse.

Le chapitre 2 s'intéresse à la qualité du M -estimateur de φ . Cette étude utilise les développements de McNeney et Wellner (2000) sur la théorie des estimateurs efficaces pour des modèles semi-paramétriques où les variables ne sont pas indépendantes et identiquement distribuées. Les principaux résultats et définitions sont brièvement rappelés dans la section 1.2.

Enfin, le chapitre 3 présente deux procédures de test d'adéquation pour les modèles étudiés dans le chapitre 2. L'une de ces procédures nécessite un algorithme bootstrap pour sa mise en oeuvre. La section 1.3 donne les principes généraux du Bootstrap.

1.1 M -estimation

Avec les notations introduites dans l'introduction, cette section reprend la terminologie utilisée par Dacunha-Castelle et Duflo (1986) au chapitre 3, ainsi que les résultats présentés par Van der Vaart (1998) au chapitre 5. Les estimateurs du paramètre $\varphi \in \Theta_\varphi \subset \mathbb{R}^d$, qui sont définis par la suite, sont des M -estimateurs : ils minimisent ou maximisent des fonctions définies à partir des données, appelées processus de contraste.

Définition 1.1. *On considère un modèle statistique semi-paramétrique $(\mathcal{X}, \mathcal{T}, \mathcal{P})$. On appelle fonction de contraste de ce modèle relativement à φ^* une fonction $M : \varphi \in \Theta_\varphi \rightarrow M(\varphi) \in \mathbb{R}_+$ ayant un minimum (resp. un maximum) strict en φ^* .*

Si les expériences sont décrites par la filtration (\mathcal{T}_n) , un processus de contraste relatif à φ et à M est une suite de fonctions $M_n : \Theta_\varphi \times \Omega \rightarrow \mathbb{R}_+$ telle que :

- $\forall (\varphi, n) \in \Theta_\varphi \times \mathbb{N}$, $M_n(\varphi) : \Omega \rightarrow \mathbb{R}_+$ est \mathcal{T}_n -mesurable.

- $M_n(\varphi)$ converge en probabilité sous $\mathbb{P}_{\varphi^*, \eta^*}$ vers $M(\varphi)$.

Enfin, un estimateur de minimum (resp. maximum) de contraste associé à M_n est un estimateur \mathcal{T}_n -adapté ($\hat{\varphi}_n$) tel que,

$$\hat{\varphi}_n = \arg \min_{\varphi \in \Theta_\varphi} M_n(\varphi), \quad (\text{resp. } \arg \max.)$$

Des exemples bien connus d'estimateurs du maximum ou du minimum de contraste sont les estimateurs du maximum de vraisemblance ou encore les estimateurs des moindres carrés.

Une première qualité recherchée sur les estimateurs est la consistance, c'est-à-dire ($\hat{\varphi}_n$) converge en probabilité vers φ^* sous $\mathbb{P}_{\varphi^*, \eta^*}^{(n)}$. [Dacunha-Castelle et Duflo \(1986\)](#) énonce un théorème de consistance basé sur la convergence du module de continuité. Ici, nous utilisons le théorème présenté par [Van der Vaart \(1998\)](#) basé sur l'uniforme convergence en probabilité du processus de contraste.

Théorème 1.1 (Consistance). *Soit (M_n) une suite de processus de contraste relatif à φ^* et à une fonction de contraste M . On note $\mathbb{P} = \mathbb{P}_{\varphi^*, \eta^*}$. Si M_n converge uniformément en probabilité vers M sur Θ_ϕ ,*

$$\sup_{\varphi \in \Theta_\varphi} |M_n(\varphi) - M(\varphi)| = o_{\mathbb{P}}(1),$$

Alors toute suite d'estimateurs $(\hat{\varphi}_n)$ vérifiant $M_n(\hat{\varphi}_n) \leq M_n(\varphi_0) + o_{\mathbb{P}}(1)$ converge en probabilité vers φ_0 .

Le théorème 1.1 est utilisé en supposant que l'espace des paramètres Θ_ϕ est compact. Cette hypothèse dans la pratique est bien souvent nécessaire pour pouvoir établir l'uniforme convergence en probabilité de M_n vers M . Elle peut sembler au premier abord peu naturelle, notamment pour les modèles étudiés aux chapitres 4 et 5. Mais on peut considérer cette hypothèse de compacité comme une première approximation grossière permettant de limiter le domaine d'approximation.

Supposons que la suite d'estimateurs $(\hat{\varphi}_n)$ construite soit consistante pour φ^* . On étudie alors sa vitesse de convergence. Plus précisément, on recherche une suite (R_n) de réels, qui tend vers $+\infty$ avec n , de telle sorte que la suite des fluctuations $(R_n(\hat{\varphi}_n - \varphi^*))_n$ reste bornée, c'est-à-dire que la suite de variables $(R_n(\hat{\varphi}_n - \varphi^*))_n$ soit tendue :

$$\forall \epsilon > 0, \exists K_\epsilon, \quad \text{tel que} \quad \mathbb{P}_{\varphi^*, \eta^*}(\|R_n(\hat{\varphi}_n - \varphi^*)\| \leq K_\epsilon) \geq 1 - \epsilon.$$

On dit alors que $(\hat{\varphi})_n$ est R_n -consistant pour φ^* . Dans certains modèles, il est possible de montrer que la suite $R_n(\hat{\varphi}_n - \varphi^*)$ converge en loi vers une loi $\mathcal{L}_{\vartheta^*}$ sous \mathbb{P}_{ϑ^*} . Cette

loi limite permet alors de construire des intervalles de confiance pour φ^* . Quand $\mathcal{L}_{\vartheta^*}$ est une loi gaussienne, on dit que $\hat{\varphi}_n$ est asymptotiquement normale.

Lorsque le processus de contraste est deux fois différentiable selon le paramètre φ , alors $\hat{\varphi}$ est un Z -estimateur pour φ^* , c'est-à-dire qu'il est solution de l'équation,

$$\nabla M_n(\varphi) = 0,$$

où ∇ est l'opérateur gradient. Par un développement limité au voisinage de φ^* du gradient, il existe $\tilde{\varphi}_n$ appartenant à l'intervalle $[\hat{\varphi}_n, \varphi^*]$ tel que,

$$0 = \nabla M_n(\hat{\varphi}_n) = \nabla M_n(\varphi^*) + \nabla^2 M_n(\tilde{\varphi}_n),$$

où ∇^2 désigne l'opérateur hessien. Cette dernière équation est utilisée pour déterminer la vitesse de convergence de l'estimateur. En effet, il suffit de rechercher la vitesse R_n à laquelle $R_n \nabla M_n(\varphi^*)$ est tendue, et de vérifier que $\nabla^2 M_n(\tilde{\varphi}_n)$ converge en probabilité vers une matrice inversible.

1.2 Efficacité dans des modèles semi-paramétriques

Parmi tous les estimateurs de φ^* que l'on peut définir, on recherche l'estimateur le plus pertinent. La précision d'un estimateur T_n est mesurée par l'erreur quadratique moyenne :

$$\mathbb{E} [(T_n - \varphi)^2],$$

où \mathbb{E} est l'espérance mathématique associée à $\mathbb{P} = \mathbb{P}_{\varphi^*, \eta^*}$. Pour rendre l'erreur quadratique moyenne la plus petite, il faut que l'estimateur soit sans biais et que la variance soit petite. Parmi les estimateurs sans biais, on choisit donc celui qui a la variance la plus petite, cette propriété traduit l'*efficacité* de l'estimateur. Dans l'étude des modèles paramétriques $\Theta_\eta = \{\eta^*\}$, il est bien connu que, sous certaines conditions de régularité du modèle, cette variance ne peut descendre en dessous d'un certain seuil, appelée borne de Cramer Rao :

$$\text{var}(\sqrt{n}T_n) \geq I_{\varphi^*, \eta^*}^{-1},$$

où I_{φ^*, η^*} est l'information de Fisher. Cette borne n'est généralement pas atteinte à n fixé, mais seulement asymptotiquement. On parle alors d'un estimateur asymptotiquement efficace. Afin d'éviter les phénomènes de super-efficacité (voir par exemple l'exemple de l'estimateur de Hodges) qui peut éventuellement se produire pour un ensemble dénombrable de φ^* , il est nécessaire non seulement d'étudier l'estimateur ponctuellement pour tout $\varphi^* \in \Theta_\varphi$, mais aussi aux voisinages locaux $(\varphi^* + h/\sqrt{n})$,

$h \in \mathbb{R}^d$. Nous renvoyons au chapitre 8 de [Van der Vaart \(1998\)](#) pour une étude détaillée de l'efficacité dans des modèles paramétriques.

Depuis une vingtaine d'années, la notion d'estimation efficace dans des modèles semi-paramétriques a été développée, d'abord pour des variables indépendantes identiquement distribuées (i.i.d.) ([Bickel et al. \(1993\)](#), [Van der Vaart \(1998, 2002\)](#)), et plus récemment pour des variables non i.i.d. ([Strasser \(1989\)](#), [McNeney et Wellner \(2000\)](#), [Hallin et Werker \(2003\)](#)). Les modèles que nous considérons dans cette thèse sont des modèles semi-paramétriques où les observations sont indépendantes non identiquement distribuées. L'étude sur l'efficacité du chapitre 2 utilise les résultats établis par [McNeney et Wellner \(2000\)](#). Ils sont rappelés dans la sous-section suivante.

1.2.1 Estimation semi-paramétrique efficace

Information du modèle. La définition de l'information dans des modèles semi-paramétriques part de l'idée suivante : estimer le paramètre $\nu(\mathbb{P})$ à partir d'un modèle \mathcal{P} est certainement plus difficile que d'estimer ce même paramètre dans un sous-modèle $\mathcal{P}_0 \subset \mathcal{P}$. L'approche semi-paramétrique considère donc le plus souvent des sous-modèles paramétriques pour lesquels est définie l'information de Fisher. Le problème est alors de savoir quels sous-modèles doivent être considérés pour ne pas s'éloigner trop du modèle semi-paramétrique de départ. Puisque l'idée que l'on se fait de "l'information" pour estimer $\nu(\mathbb{P})$ dans le modèle entier est qu'elle est certainement plus petite que la borne inférieure sur tous les sous-modèles paramétriques, on est donc intéressé par les suites de sous-modèles qui atteignent asymptotiquement cette borne. Un sous-modèle qui atteint cette borne inférieure est appelé le *sous-modèle le moins favorable*.

La propriété LAN. Comme pour les modèles paramétriques, il faut poser des hypothèses de régularité sur le modèle. De telles hypothèses se généralisent par la propriété de Normalité Asymptotique Locale (LAN). On note le log-rapport de vraisemblance entre deux points ϑ_1 et ϑ_2 de Θ pour n observations de la façon suivante :

$$\Lambda_n(\vartheta_1, \vartheta_2) = \log \frac{d\mathbb{P}_{\vartheta_1}^{(n)}}{d\mathbb{P}_{\vartheta_2}^{(n)}},$$

où $d\mathbb{P}_{\vartheta_1}^{(n)}/d\mathbb{P}_{\vartheta_2}^{(n)}$ est la dérivée de Radon-Nikodym de la partie absolument continue de $\mathbb{P}_{\vartheta_1}^{(n)}$ par rapport à $\mathbb{P}_{\vartheta_2}^{(n)}$.

Définition 1.2. Soit \mathcal{H} un espace vectoriel muni d'un produit scalaire noté $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ (de norme associée $\|\cdot\|_{\mathcal{H}}$). Le modèle $\mathcal{P}^{(n)}$ est Localement Asymptotiquement Normal

(LAN) en $\vartheta^* \in \Theta$ indexé par l'espace tangent \mathcal{H} si pour tout $h \in \mathcal{H}$ il existe une suite $\mathbb{P}_{\vartheta_n(h)}$ de mesure de probabilité telle que

$$\Lambda_n(\vartheta_n(h), \vartheta^*) = \Delta_n(h) - \frac{1}{2} \|h\|_{\mathcal{H}}^2 + o_{\mathbb{P}_{\vartheta^*}^{(n)}}(1).$$

$\Delta_n(h)$ est un processus aléatoire centré qui vérifie

$$\mathcal{L}(\Delta_n(h_1), \dots, \Delta_n(h_d) | \mathbb{P}_{\vartheta^*}^{(n)}) \rightarrow \mathcal{N}_d(0, (< h_i, h_j >_{\mathcal{H}})_{1 \leq i, j \leq d}),$$

pour toute sous-famille finie $(h_1, \dots, h_d) \in \mathcal{H}^d$.

L'exemple suivant illustre la propriété LAN pour des modèles semi-paramétriques où les observations sont i.i.d.

Exemple 1

Soient $X_1 \dots X_n$ un échantillon de loi $\mathbb{P} \in \mathcal{P} = \{\mathbb{P}_{\vartheta}, \vartheta \in \Theta\}$. On suppose que tout sous-modèle de dimension un, $(\mathbb{P}_t)_{0 \leq t < \epsilon}$, est différentiable en moyenne quadratique en $t = 0$. C'est-à-dire il existe une fonction $g \in \mathbb{L}^2(\mathbb{P}_0)$, appelée *fonction score*, telle que

$$\int g d\mathbb{P}_0 = 0, \quad \text{et} \quad \int \left[\frac{d\mathbb{P}_t^{1/2} - d\mathbb{P}_0^{1/2}}{t} - \frac{1}{2} g d\mathbb{P}_0^{1/2} \right]^2 \rightarrow 0.$$

Le log-rapport de vraisemblance entre $t = 0$ et $t = 1/\sqrt{n}$ est :

$$\log \prod_{i=1}^n \frac{d\mathbb{P}_{1/\sqrt{n}}}{d\mathbb{P}_0}(X_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^n g(X_i) - \frac{1}{2} \mathbb{P} g^2 + o_{\mathbb{P}}(1).$$

Ce modèle est donc LAN, et l'espace tangent associé est l'ensemble des fonctions scores g .

Plus particulièrement, si \mathbb{P}_{ϑ} admet une densité p_{ϑ} (avec $\vartheta = (\varphi, \eta)$), un sous-modèle de dimension un est du type $(\mathbb{P}_{\varphi^* + ta, \eta_t})_t$ avec $a \in \mathbb{R}^d$ et $t \rightarrow \eta_t$ un chemin dans Θ_{η} . Si ces sous-modèles sont différentiables en moyenne quadratique en $t = 0$, l'espace tangent associé est :

$$\mathcal{H} = \{a \cdot \nabla l_{\varphi^*, \eta^*} + g, \ a \in \mathbb{R}^d, g \in {}_{\eta} \mathcal{H}_{\varphi^*, \eta^*}\},$$

où $l_{\varphi^*, \eta^*} = \log p_{\varphi^*, \eta^*}$, ∇ est l'opérateur gradient par rapport au paramètre fini-dimensionnel φ , et ${}_{\eta} \mathcal{H}_{\varphi^*, \eta^*}$ est un sous-espace de $\{g \in \mathbb{L}^2(\mathbb{P}_{\varphi^*, \eta^*}), \ \mathbb{P}_{\varphi^*, \eta^*} g = 0\}$.

Comme nous l'avons mentionné plus haut, pour étudier l'efficacité de l'estimation on considère les voisinages de $\vartheta^* = (\varphi^*, \eta^*)$ du type $\vartheta + h/\sqrt{n}$ avec $h \in \mathcal{H}$. On doit alors supposer que le paramètre à estimer $\nu_n(\mathbb{P}_{\vartheta^*})$ est suffisamment régulier, c'est-à-dire différentiable par rapport à l'espace tangent \mathcal{H} . Dans l'optique de l'étude de

ces voisinages locaux de ϑ^* , on est donc particulièrement intéressé par les estimateurs atteignant localement uniformément leur distribution limite : les estimateurs réguliers. Pour ces estimateurs, un petit changement dans le paramètre ϑ^* affecte peu la loi limite, et un changement négligeable ne modifie pas la loi de l'estimateur.

Définition 1.3. Supposons que $\nu_n(\mathbb{P}_{\vartheta}^{(n)})$ soit à valeur dans un espace de Banach \mathbb{B} . On dit la suite (ν_n) est différentiable en ϑ relativement à \mathcal{H} s'il existe un opérateur continue $\dot{\nu} : \mathcal{H} \rightarrow \mathbb{B}$ et une suite d'opérateurs linéaires continus $R_n : \mathbb{B} \rightarrow \mathbb{B}$ tel que $\|R_n\| \rightarrow \infty$ et

$$R_n \left(\nu_n(\mathbb{P}_{\vartheta_n(h)}^{(n)}) - \nu_n(\mathbb{P}_{\vartheta^*}^{(n)}) \right) \rightarrow \dot{\nu}(h), \quad \text{pour tout } h \in \mathcal{H}.$$

On note alors $\tilde{\nu}$ le représentant de $\dot{\nu}$ dans le complété $\text{cl}(\mathcal{H})$ de \mathcal{H} .

Exemple 1 (suite)

Le paramètre $\nu_n(\mathbb{P}_{\varphi^*, \eta^*}) = \varphi^*$ est différentiable en φ_0 relativement à \mathcal{H} avec $R_n = \sqrt{n}$. Le représentant de la différentielle dans la fermeture de l'espace tangent est $\tilde{I}_{\varphi^*, \eta^*}^{-1} \tilde{l}_{\varphi^*, \eta^*}$, appelé fonction d'influence efficace. La fonction $\tilde{l}_{\varphi^*, \eta^*} = \nabla l_{\varphi^*, \eta^*} - \Pi_{\varphi^*, \eta^*}(\nabla l_{\varphi^*, \eta^*})$ est appelée *score efficace*, avec Π_{φ^*, η^*} l'opérateur de projection orthogonale sur la fermeture de ${}_{\eta}\mathcal{H}_{\varphi^*, \eta^*}$. La matrice carrée $\tilde{I}_{\varphi^*, \eta^*}^{-1} = \mathbb{P}_{\varphi^*, \eta^*} \left(\tilde{l}_{\varphi^*, \eta^*} \tilde{l}_{\varphi^*, \eta^*}^T \right)$ est appelée *matrice d'information efficace*.

Définition 1.4. Soit $\mathcal{P}^{(n)}$ un modèle LAN en ϑ^* indexé par l'espace tangent \mathcal{H} . Soit T_n un estimateur du paramètre ν_n différentiable en ϑ^* . T_n est dit régulier si pour tout $h \in \mathcal{H}$, $R_n(T_n - \nu_n(\mathbb{P}_{\vartheta_n(h)}^{(n)}))$ converge en loi sous $\mathbb{P}_{\vartheta_n(h)}^{(n)}$ vers une distribution L_{ϑ^*} définie sur \mathbb{R} qui ne dépend pas de $h \in \mathcal{H}$.

La proposition 1.1 met en relation les estimateurs asymptotiquement linéaires et les estimateurs asymptotiquement réguliers efficaces.

Définition 1.5. Soit $\mathcal{P}^{(n)}$ un modèle LAN en ϑ^* indexé par l'espace tangent \mathcal{H} . Soit T_n un "estimateur du paramètre ν_n . T_n est dit asymptotiquement linéaire si pour toute forme linéaire $b^* : \mathbb{B} \rightarrow \mathbb{R}$ il existe $h_{b^*} \in \mathcal{H}$ tel que

$$\sqrt{n}b^*(T_n - \nu_n(\mathbb{P}_{\vartheta^*}^{(n)})) = \Delta_n(h_{b^*}) + o_{\mathbb{P}}(1).$$

En particulier si $\nu_n(\mathbb{P}_{\vartheta^*}^{(n)})$ est d -dimensionnel, T_n est asymptotiquement linéaire si, et seulement si :

$$\sqrt{n}(T_n - \nu_n(\mathbb{P}_{\vartheta^*}^{(n)})) = \Delta_n(\tilde{h}) + o_{\mathbb{P}}(1),$$

où $\tilde{h} = (h_1, \dots, h_d) \in \mathcal{H}^d$ et $\Delta_n(\tilde{h}) = (\Delta_n(h_1), \dots, \Delta_n(h_d))$

La proposition suivante établie caractérise les estimateurs asymptotiquement efficaces dans le cas fini-dimensionnel, tandis que la proposition 1.2 la généralise pour des espaces de Banach.

Proposition 1.1 (McNeney et Wellner (2000)). *Soit $\mathcal{P}^{(n)}$ un modèle LAN en ϑ^* indexé par l'espace tangent \mathcal{H} tel que le processus centré $\Delta_n(h)$ soit linéaire en h . Soit $\nu_n(\mathbb{P}_{\vartheta}^{(n)})$ un paramètre d -différentiable relativement à l'espace \mathcal{H} . On note $\dot{\nu} \in \overline{\mathcal{H}}^d$ le représentant de la différentielle de ν_n . Soit T_n un estimateur asymptotiquement linéaire en $\Delta_n(\tilde{h})$. Alors T_n est régulier efficace si, et seulement si $\tilde{h} = \dot{\nu}$.*

Exemple 1 (suite)

Par conséquent un estimateur T_n de φ est asymptotiquement efficace si,

$$\sqrt{n}(T_n - \phi_0) = \left(\tilde{I}_{\varphi_0, \eta_0} \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{l}_{\varphi_0, \eta_0}(X_i) \right) + o_{\mathbb{P}}(1).$$

Proposition 1.2 (McNeney et Wellner (2000)). *Soit $\mathcal{P}^{(n)}$ un modèle LAN en ϑ^* indexé par l'espace tangent \mathcal{H} tel que le processus centré $\Delta_n(h)$ soit linéaire en h . Soit $\nu_n(\mathbb{P}_{\vartheta}^{(n)})$ une suite de paramètres à valeur dans un espace normé complet \mathbb{B} . On suppose que cette suite est différentiable de dérivée $\dot{\nu}$, telle que pour toute forme linéaire b^* de \mathbb{B} , on note $\dot{\nu}_{b^*} \in \text{cl}(\mathcal{H})$ le représentant de $b^*\dot{\nu}$.*

Si T_n est asymptotiquement linéaire au sens de la définition 1.5 avec $h_{b^} \in \mathcal{H}$, alors (b^*T_n) est régulier efficace si, et seulement si $h_{b^*} = \dot{\nu}_{b^*}$. De plus, si $\sqrt{n}(T_n - \nu_n(\mathbb{P}_{\vartheta_n(h)}^{(n)}))$ converge faiblement sous $(\mathbb{P}_{\vartheta_n(h)}^{(n)})$ vers une limite tendue de \mathbb{B} pour tout suite $\vartheta_n(h)$, alors T_n est régulier efficace.*

1.2.2 Vraisemblance profilée

Dans les modèles paramétriques, une méthode connue pour obtenir un estimateur efficace est de maximiser la vraisemblance du modèle. Or une telle méthode est difficile à mettre en oeuvre pour des modèles de grande dimension, en particulier pour les modèles semi-paramétriques. Dans ces derniers modèles, l'information observée, si elle existe, serait au mieux un opérateur infini-dimensionnel. Souvent ce problème est surmonté en employant une vraisemblance profilée.

Définition 1.6. *Supposons que le paramètre est divisé comme $\vartheta = (\varphi, \eta)$, avec φ un paramètre fini-dimensionnel et η un paramètre de nuisance. Notons $l_n(\vartheta, \eta)$ la vraisemblance du modèle, alors la vraisemblance profilée se définit comme*

$$pl_n(\vartheta) = \sup_{\eta \in \Theta_{\eta}} l_n(\vartheta, \eta).$$

La vraisemblance profilée consiste donc à maximiser la vraisemblance en deux étapes, d'abord par rapport à la nuisance, ensuite par rapport au paramètre fini-dimensionnel.

Soit $(\hat{\varphi}_n, \hat{\eta}_n)$ l'estimateur du maximum de vraisemblance de (φ^*, η^*) , alors $\hat{\varphi}_n$ maximise la vraisemblance profilée $\varphi \rightarrow pl_n(\varphi)$. Pour des modèles semi-paramétriques indépendants identiquement distribués, [Murphy et Van der Vaart \(2000\)](#) ont montré que sous certaines conditions de régularité sur le modèle (similaires à celles qui assurent que l'estimateur du maximum de vraisemblance soit efficace) la vraisemblance profilée admet le développement suivant pour toute suite $\tilde{\varphi}_n$ convergente en probabilité vers φ^* :

$$\begin{aligned} \log pl_n(\tilde{\varphi}_n) = \log pl_n(\varphi^*) &+ (\tilde{\varphi}_n - \varphi^*)^T \sum_{i=1}^n \tilde{l}_{\varphi^*, \eta^*}(X_{i,n}) \\ &+ \frac{1}{2}(\tilde{\varphi}_n - \varphi^*)^T \tilde{I}_{\varphi^*, \eta^*}(\tilde{\varphi}_n - \varphi^*) + o_{\mathbb{P}_{\varphi^*, \eta^*}^{(n)}}(\sqrt{n}\|\tilde{\varphi}_n - \varphi^*\| + 1). \end{aligned} \quad (1.1)$$

Ainsi, si $\tilde{I}_{\varphi_0, \eta_0}$ est inversible et $\hat{\varphi}_n$ est consistant, $\hat{\varphi}_n$ est asymptotiquement efficace.

Dans le chapitre 2, nous utilisons un processus de contraste construit à partir de la vraisemblance profilée. L'estimateur qui maximise ce contraste est, sous certaines conditions de régularité du modèle, asymptotiquement efficace. D'autre part la vraisemblance profilée du modèle admet un développement asymptotique similaire à (1.1).

1.3 Méthode bootstrap

Le bootstrap est une technique inventée par [Efron \(1979\)](#). Elle s'appuie sur le fait de pouvoir estimer, en re-échantillonnant dans les données, les caractéristiques du phénomène aléatoire qui a engendré ces données. La technique du bootstrap mime donc la distribution d'un estimateur ou d'une statistique de test afin de trouver des approximations de quantités difficilement calculables. Dans le chapitre 3, nous construisons des tests statistiques pour valider une sous-classe du modèle de forme invariante pour des données. Or sous l'hypothèse nulle, la loi limite de la statistique de test est trop compliquée pour utiliser le principe du "plug-in" afin d'estimer les quantiles. On utilise donc le bootstrap afin d'en avoir une approximation. Cette section a pour but de donner les grands principes de cette méthode et de l'illustrer avec des exemples simples.

Soit $\underline{X} = (X_1, \dots, X_n)$ un échantillon d'une loi \mathbb{F} . On note par $\underline{x} = (x_1, \dots, x_n)$ les valeurs effectivement mesurées. On souhaite estimer un paramètre $\theta = t(\mathbb{F})$ à partir de \underline{x} . Si n est assez grand, la distribution empirique de l'échantillon $\hat{\mathbb{F}}$ est proche de \mathbb{F}

(théorème de Glivenko-Cantelli). Le principe du "plug-in" consiste alors à considérer l'estimateur suivant $\hat{\theta} = t(\hat{\mathbb{F}})$.

Le bootstrap consiste à générer B échantillon bootstrap à partir de \underline{x} . Un *échantillon bootstrap* \underline{x}^* est obtenu en tirant aléatoirement les observations de l'échantillon de départ \underline{x} , autrement dit à effectuer des tirages avec remise des n valeurs observées. En terme de variables aléatoires, on peut résumer de la façon suivante :

ÉCHANTILLON	ÉCHANTILLON BOOTSTRAP
$\underline{X} = (X_1, \dots, X_n)$ généré par \mathbb{F} , de loi empirique $\hat{\mathbb{F}}$	$\underline{X}^* = (X_1^*, \dots, X_n^*)$ généré par $\hat{\mathbb{F}}$, de loi empirique $\hat{\mathbb{F}}^*$

Pour chaque échantillon bootstrap \underline{x}_b^* , $b = 1, \dots, B$, on note $\hat{\mathbb{F}}_b^*$ la distribution empirique et on calcule la statistique $\hat{\theta}_b^* = t(\hat{\mathbb{F}}_b^*)$. Le principe du bootstrap est de considérer la moyenne $\hat{\theta}_{boot}$ des valeurs $\hat{\theta}_b^*$, $b = 1, \dots, B$, pour approcher la valeur $\theta = t(\mathbb{F})$,

$$\hat{\theta}_{boot} = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^*.$$

Pour certains modèles, il a été prouvé, que pour un B assez grand, la statistique bootstrap $\hat{\theta}_{boot}$ pouvait être une meilleure approximation de $\hat{\theta}$ (voir [Hall \(1992\)](#)).

Exemple 2 (Algorithme bootstrap pour estimer les écarts-type)

1. On génère B échantillons bootstrap indépendants \underline{x}_b^* , $b = 1, \dots, B$.
2. On évalue la réplique bootstrap correspondant à chaque échantillon bootstrap : $\hat{\theta}_b^*$, $b = 1, \dots, B$.
3. On estime l'écart-type de $\hat{\theta}$ par l'écart type des B répliques,

$$\left\{ \frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_b^* - \hat{\theta}_{boot})^2 \right\}^{1/2}.$$

Plus précisément, supposons une équation dite "équation de population" qui relie la population \mathbb{F} à l'échantillon $\hat{\mathbb{F}}$,

$$\mathbb{E}(f_\theta(\mathbb{F}, \hat{\mathbb{F}})/\mathbb{F}) = 0, \quad \theta \in \mathbb{R}.$$

On cherche à résoudre cette équation selon la variable θ . On note θ_0 la solution de cette équation.

Exemple 3 (Equation de population pour la variance)

Soit $\sigma^2 = \text{var}(X_1)$. Alors σ^2 est solution de l'équation de population,

$$\begin{cases} \mathbb{E}(f_\theta(\mathbb{F}, \hat{\mathbb{F}})/\mathbb{F}) = 0, & \theta \in \mathbb{R}_+, \\ f_\theta(\mathbb{F}, \hat{\mathbb{F}}) = (\int x d\hat{\mathbb{F}} - \int x d\mathbb{F})^2 - \theta. \end{cases}$$

Exemple 4 (Estimation d'une quantile)

Soit $T = T(\underline{X})$ une statistique fonction de l'échantillon. La loi de T est donc une fonction de \mathbb{F} . On souhaite estimer la quantile t_α d'ordre $\alpha \in [0, 1]$ de cette statistique. t_α est solution de l'équation de population,

$$\begin{cases} \mathbb{E}(f_\theta(\mathbb{F})/\mathbb{F}) = 0, & \theta \in \mathbb{R}, \\ f_\theta(\mathbb{F}) = 1_{(T \leq \theta)} - \alpha. \end{cases}$$

Le principe du bootstrap consiste à remplacer le couple $(\mathbb{F}, \hat{\mathbb{F}})$ par $(\hat{\mathbb{F}}, \hat{\mathbb{F}}^*)$ dans l'équation de population afin d'approcher la solution t_α . Ceci repose sur l'idée qu'il existe la même relation entre \mathbb{F} et $\hat{\mathbb{F}}$, qu'entre $\hat{\mathbb{F}}$ et $\hat{\mathbb{F}}^*$. Soit $\hat{\theta}_0$ la solution de l'équation suivante,

$$\mathbb{E}(f_\theta(\hat{\mathbb{F}}, \hat{\mathbb{F}}^*)/\hat{\mathbb{F}}) = 0, \quad \theta \in \mathbb{R}. \quad (1.2)$$

On dit que $\hat{\theta}_0$ est l'estimateur bootstrap de θ_0 .

Exemple 3 (Equation de population pour la variance (suite))

L'estimateur bootstrap de la variance est,

$$S^2 = \mathbb{E} \left[\left(\int x d\hat{\mathbb{F}}^* - \int x d\hat{\mathbb{F}} \right)^2 / \hat{\mathbb{F}} \right] = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2,$$

où \bar{X} est la moyenne empirique des observations.

On peut alors utiliser des méthodes du type Monte-Carlo pour résoudre l'équation bootstrap de l'échantillon (1.2).

Exemple 4 (Algorithme bootstrap pour estimer les quantiles (suite))

1. Soit $B \in \mathbb{N}$ tel qu'il existe $v \in \mathbb{N}$ vérifiant,

$$v \leq B, \quad \text{et} \quad \frac{v}{B+1} = \alpha.$$

On génère B échantillons bootstrap indépendants : \underline{X}_b^* , $b = 1, \dots, B$.

2. Pour tout $b = 1, \dots, B$, on évalue $T^*(b) = T(\underline{X}_b^*)$.

3. $\hat{\theta}_{boot}$ est la $v^{\text{ème}}$ plus grande valeur de $\{T^*(b) = T(\underline{X}_b^*), b = 1, \dots, B\}$.

$\hat{\theta}_{boot}$ est un estimateur de θ_0 , et quand B tend vers l'infini, $\hat{\theta}_{boot}$ converge vers θ_0 .

1.4 Transformée de Fourier

Le plus souvent, l'analyse de Fourier est entendue comme l'étude des séries de Fourier (cercle unité), ou encore l'étude des transformées de Fourier sur la droite

réelle. Durant les années 1950, les concepts de base et les théorèmes ont été généralisés aux groupes abéliens localement compacts. Dans cette thèse, nous considérons quatre types de transformées de Fourier : sur le tore (séries de Fourier) dans le chapitre 2, sur le plan réel (transformée de Fourier classique dans \mathbb{R}^2) et sur le plan polaire (transformée de Fourier-Mellin) dans le chapitre 4, et sur les entiers relatifs dans le chapitre 5. Cette section a pour but de connecter entre elles ces quatre transformées, et de présenter dans les grandes lignes la théorie de Fourier sur les groupes localement compacts. Le lecteur peut se reporter aux ouvrages de Rudin (1990) et Ramakrishnan et Valenza (1999) pour un exposé détaillé.

Groupe topologique et mesure de Haar

Un groupe topologique abélien est un groupe $(G, +)$ commutatif, muni d'une topologie telle que l'application $(x, y) \rightarrow x - y$ de l'espace $G \times G$ dans G soit continue. Si de plus, la topologie de G est localement compacte, on dit que G est un *groupe abélien localement compact* (LCA).

Exemple 5

Quelques exemples de groupes abéliens localement compacts,

1. tout groupe au plus dénombrable muni de la topologie discrète, par exemple $(\mathbb{Z}, +)$;
2. (\mathbb{R}^*, \times) , (\mathbb{R}_+^*, \times) et (\mathbb{C}^*, \times) muni de la topologie induite ;
3. tout groupe additif d'un espace vectoriel de dimension fini, comme $(\mathbb{R}^d, +)$, ou $(\mathbb{C}^d, +)$;
4. $(\mathbb{T}^d, +)$ ($\mathbb{T} = \mathbb{R}/(2\pi\mathbb{Z})$ est le tore) muni de la topologie quotient ;

Pour tout groupe abélien localement compact G , on appelle *mesure de Haar*, une mesure de Radon μ (finie pour tout sous ensemble compact de G , extérieurement régulière) invariante par translation, c'est-à-dire,

$$\mu(E + x) = \mu(E),$$

pour tout $x \in G$ et pour tout borélien E de G . Le théorème de Haar montre l'existence et l'unicité (à une constante multiplicative près) de la mesure de Haar pour un groupe localement compact.

Exemple 6

Quelques exemples de groupes abéliens localement compacts associés à leur mesure de Haar,

1. la mesure de comptage de $(\mathbb{Z}, +)$;
2. la mesure de Lebesgue de $(\mathbb{R}^d, +)$, $(\mathbb{T}^d, +)$;
3. la mesure $d\mu(r) = r^{-1}dr$ pour (\mathbb{R}_+^*, \times) .

Dualité de Pontrjagin et transformée de Fourier

Soit G un groupe LCA. Une fonction $\gamma : G \rightarrow \mathbb{C}^*$ est un *caractère* de G si c'est un homomorphisme du groupe G dans le groupe multiplicatif \mathbb{C}^* tel que,

$$|\gamma(x)| = 1, \quad \text{et} \quad \gamma(x+y) = \gamma(x)\gamma(y), \quad \forall (x, y) \in G^2.$$

L'ensemble \hat{G} des caractères de G est un groupe abélien pour l'opération interne suivante,

$$(\gamma_1 + \gamma_2)(x) = \gamma_1(x)\gamma_2(x), \quad x \in G, \quad (\gamma_1, \gamma_2) \in \hat{G}.$$

On appelle \hat{G} *le groupe dual de Pontrjagin* de G . On muni \hat{G} d'une topologie de telle sorte que l'application $(x, \gamma) \rightarrow \gamma(x)$ de $G \times \hat{G}$ dans \mathbb{C} soit continue. On montre alors que

- si G est LCA, alors \hat{G} est LCA ;
- si G est compact, alors \hat{G} est discret ;
- si G est discret, alors \hat{G} est compact.

Notons $\mathbb{L}^1(G, \mu)$ l'espace des fonctions intégrables à valeurs complexes pour la mesure de Haar μ associée à G .

Définition 1.7. Soit $f \in \mathbb{L}^1(G)$. On définit la transformée de Fourier de f par,

$$\hat{f}(\gamma) = \int_G f(x)\bar{\gamma}(x)d\mu(x), \quad \gamma \in \hat{G}.$$

Exemple 7 (Quelques transformations de Fourier) 1. si $G = \mathbb{R}$, le groupe dual associé est l'ensemble des fonctions

$$x \in \mathbb{R} \rightarrow e^{i\omega x}, \quad \omega \in \mathbb{R}.$$

On peut donc identifier \hat{G} au groupe additif \mathbb{R} . Dans ce cas, la définition 1.7 est la transformée de Fourier classique pour $f \in \mathbb{L}^1(\mathbb{R}, dx/\sqrt{2\pi})$,

$$\hat{f}(\omega) = \int_{\mathbb{R}} f(x)e^{-i\omega x} \frac{dx}{\sqrt{2\pi}}, \quad \omega \in \mathbb{R};$$

2. si $G = \mathbb{R}_+^*$, le groupe dual associé est l'ensemble des fonctions

$$r \in \mathbb{R}_+^* \rightarrow r^{iv}, \quad v \in \mathbb{R}.$$

On peut donc identifier \hat{G} au groupe additif \mathbb{R} . Dans ce cas, la définition 1.7 est la transformée de Mellin pour $f \in \mathbb{L}^1(\mathbb{R}_+^*, dr/r)$,

$$\mathcal{M}_f(v) = \int_0^{+\infty} f(r) r^{-iv} \frac{dr}{r}, \quad v \in \mathbb{R};$$

3. si $G = \mathbb{T}$, le groupe dual associé est l'ensemble des fonctions

$$x \in \mathbb{R} \rightarrow e^{ikx}, \quad k \in \mathbb{Z}.$$

On peut donc identifier \hat{G} au groupe additif \mathbb{Z} . Dans ce cas, la définition 1.7 sont les coefficients de Fourier de $f \in \mathbb{L}^1(\mathbb{T}, dx/(2\pi))$,

$$c_k(f) = \int_0^{2\pi} f(x) e^{-ikx} \frac{dx}{2\pi}, \quad k \in \mathbb{Z};$$

4. si $G = \mathbb{Z}$, le groupe dual associé est l'ensemble des fonctions

$$x \in \mathbb{R} \rightarrow e^{i\omega x}, \quad \omega \in \mathbb{T}.$$

On peut donc identifier \hat{G} au groupe additif \mathbb{T} . Dans ce cas, la définition 1.7 est la fonction génératrice de $f \in l^1(\mathbb{Z})$,

$$G_f(\omega) = \sum_{k \in \mathbb{Z}} f(k) e^{-i\omega k}, \quad \omega \in \mathbb{T},$$

où $l^1(\mathbb{Z})$ est l'ensemble des séries sommables.

En utilisant des approximations de l'unité et des opérateurs de convolution, on prouve que la transformée de Fourier est inversible, et qu'elle se prolonge par continuité en une isométrie.

Théorème 1.2 (Inversion de Fourier). *Soit $f \in \mathbb{L}^1(G, \mu)$, telle que $\hat{f} \in \mathbb{L}^1(\hat{G}, \mu_{\hat{G}})$. La mesure de Haar $\mu_{\hat{G}}$ de \hat{G} peut être normalisée de telle sorte que :*

$$f(x) = \int_{\hat{G}} \hat{f}(\gamma) \gamma(x) d\mu_{\hat{G}}(\gamma), \quad x \in G.$$

Théorème 1.3 (Théorème de Plancherel). *La transformée de Fourier restreinte au sous-espace $\mathbb{L}^1(G, \mu) \cap \mathbb{L}^2(G, \mu)$ muni de la norme \mathbb{L}^2 est une isométrie sur un sous-espace dense de $\mathbb{L}^2(\hat{G}, \mu_{\hat{G}})$. Ainsi, cette application peut se prolonger de manière unique sur $\mathbb{L}^2(G)$ dans $\mathbb{L}^2(\hat{G})$.*

On prouve, par le théorème de l'inversion de Fourier, que tout groupe LCA est identifiable au groupe dual de son groupe dual \hat{G} ,

$$G \simeq \hat{\hat{G}}.$$

Chapitre 2

Estimation des paramètres pour une sous-classe de modèle de forme invariante

Sommaire

2.1	Presentation du modèle	28
2.2	Définition des processus de contraste	32
2.3	Efficacité et normalité asymptotique	43
2.4	Application des méthodes d'estimation	53
2.5	Preuves de la Section 2.2	58
2.6	Preuves de la Section 2.3	70

Résumé : On observe un nombre fixé J de fonctions inconnues qui prennent leurs valeurs dans un espace multidimensionnel. Ces fonctions admettent en réalité une forme commune et diffèrent entre elles uniquement par un paramètre de translation, d'amplitude et de niveau. Les données sont modélisées sous un modèle de régression. Ce modèle semi-paramétrique est connu sous le nom de "Shape invariant model" (ou modèle de forme invariante). Alors que la forme générale du signal est inconnue, on propose ici des méthodes statistiques pour estimer les paramètres de translation, d'amplitude et de niveau. Les estimateurs définis sont au moins asymptotiquement normaux, voire asymptotiquement efficaces. Ces méthodes utilisent la transformée de Fourier et la vraisemblance profilée. On étudie également les propriétés asymptotiques de l'estimateur de la forme commune induit par ces méthodes.

Mots clés : Shape invariant model, estimation semi-paramétrique, efficacité semi-paramétrique, transformée de Fourier discrète.

Introduction

Dans de nombreuses études, les chercheurs étudient des observations qui peuvent dépendre de plusieurs variables. On peut penser par exemple à l'évolution de la température au cours du temps (une dimension), les niveaux de gris d'une image 2D (deux dimensions). L'objet d'intérêt n'est alors pas une variable aléatoire mais une fonction qui diffère selon le sujet étudié. Les données $(Y_{i,j})$, où l'indice j désigne le sujet étudié et l'indice i désigne le point d'observation de la mesure, sont alors modélisées de la façon suivante,

$$Y_{i,j} = f_j^*(t_{i,j}) + \sigma_j^* \epsilon_{i,j}, \quad i \in \mathcal{I}_{n,j}, \quad j = 1 \dots J. \quad (2.1)$$

C'est un modèle de régression à dépendance non linéaire : les fonctions de régression $f_j^* : \mathbb{R}^d \rightarrow \mathbb{R}$ sont inconnues et dépendent non linéairement des points d'observation $t_{i,j} \in \mathcal{D} \subset \mathbb{R}^d$, où $d \geq 1$. Les ensembles $\mathcal{I}_{n,j}$ sont des sous-ensembles finis de \mathbb{Z}^d . Les termes d'erreur $\epsilon_{i,j}$ sont des variables aléatoires indépendantes centrées de variance 1, et les réels positifs σ_j^* quantifient le niveau de bruit pour chaque sujet.

On s'intéresse au modèle où l'échantillon des fonctions de régression montre une certaine homogénéité dans la structure, dans le sens où les fonctions coïncident si elles sont proprement translatées et normalisées. En d'autres termes, il existe une fonction f^* de telle sorte que

$$f_j^*(t) = a_j^* f^*(t - \theta_j^*) + v_j^*, \quad t \in \mathbb{R}^d, \quad j = 1 \dots, J, \quad (2.2)$$

où pour tout $j = 1, \dots, J$, les paramètres $a_j^* \in \mathbb{R}$, $v_j^* \in \mathbb{R}$ et $\theta_j^* \in \mathbb{R}^d$ sont appelés respectivement paramètre d'amplitude, de niveau et de translation. Ce type de modèle a été introduit par [Lawton et al. \(1972\)](#) sous le nom de "shape invariant model" (modèle de forme invariante) : c'est en réalité une sous-classe de ce modèle. C'est un modèle semi-paramétrique. Cela signifie qu'il dépend principalement de deux paramètres, l'un fini-dimensionnel $\alpha^* = ((\theta_j^*)_j, (a_j^*)_j, (v_j^*)_j)$, l'autre infini dimensionnel f^* qui appartient à un certain grand ensemble de fonctions. Des exemples assez connus, où ce type de modèle a été considéré, est l'étude des courbes de croissances d'enfants (voir [Kneip et Gasser, 1988](#)) ou l'étude des rythmes circadiens (voir [Wang et Brown, 1996](#)).

Le principal objectif de ce chapitre est de présenter des méthodes afin d'estimer le paramètre α^* sans chercher à estimer la forme commune f^* . Le problème de l'estimation de α^* a été traité auparavant par plusieurs auteurs pour $d = 1$. D'abord [Lawton et al. \(1972\)](#) proposent une procédure statistique, SEMOR (Self-Modeling nonlinear Regression), qui estime simultanément la fonction f^* et le paramètre α^* . Cette méthode est itérative et elle est basée sur l'approximation polynomiale (des

splines d'ordre 1) de f^* sur un espace d'observation compact \mathcal{D} . La convergence et la consistance de la méthode SEMOR ont été prouvés par [Kneip et Gasser \(1988\)](#). [Härdle et Marron \(1990\)](#) étudient le modèle où les fonctions ne diffèrent que d'un paramètre de translation (θ_j^*) et de niveau (v_j^*) pour $J = 2$ sujets. Ils construisent des estimateurs consistants de vitesse de convergence \sqrt{n} et asymptotiquement normaux pour les paramètres (θ_j^*) et (v_j^*). Leur méthode est basée sur l'utilisation d'un estimateur à noyau pour estimer la fonction f^* . Par la suite, [Kneip et Engel \(1995\)](#) remplacent l'estimateur polynomial de f^* par un estimateur à noyau en reprenant la méthode SEMOR. Par ailleurs, d'autres bases de fonctions ont été choisies pour estimer la forme commune f^* , comme des splines lisses (voir [Guardabasso et al., 1988](#)) ou des splines périodiques (voir [Wang et Brown, 1996](#)). Dernièrement, pour le modèle où les fonctions f_j^* sont translatées les unes des autres, [Gamboa et al. \(2005\)](#) construisent un estimateur consistant de vitesse de convergence \sqrt{n} , asymptotiquement normal, à partir de la transformée de Fourier discrète des données. Nos méthodes d'estimation de α^* sont proches de celle présentée par [Gamboa et al. \(2005\)](#) pour des modèles plus généraux et multidimensionnels.

Ce chapitre propose trois estimateurs du paramètre α^* . Chacun de ces estimateurs est le minimum d'un processus de contraste (cf. chapitre 1) qui diffère suivant les contraintes d'identifiabilité posées sur le modèle. Ces trois jeux de contraintes d'identifiabilité sont décrites dans la section 2.1. Les processus de contraste considérés sont définis à partir de la transformée de Fourier discrète des données. En effet, cette transformation est particulièrement bien adaptée au modèle, puisqu'elle est linéaire d'une part, et d'autre part la base de vecteurs propres pour cette transformation diagonalise simultanément l'ensemble des opérateurs de translation (T_θ) sur les fonctions,

$$T_\theta : f \rightarrow f(\cdot - \theta), \quad \theta \in \mathbb{R}^d.$$

C'est pourquoi, dans ce chapitre, l'étude se limite aux cas où les fonctions f_j^* sont périodiques sur le pavé \mathcal{D} . Mais comme le souligne la remarque 3.1, cette hypothèse n'est pas aussi restrictive : les méthodes d'estimation peuvent être appliquées dans des cas où les fonctions de régression ne sont pas forcément périodiques.

Les estimateurs définis à partir de ces processus de contrastes sont tous consistants de vitesse de convergence \sqrt{n} et asymptotiquement normaux (théorème 2.4). Si les contraintes d'identifiabilité sont bien choisies, les estimateurs des paramètres de translation (θ_j^*), d'amplitude (a_j^*) et de niveau (v_j^*) sont asymptotiquement indépendants et efficaces (théorème 2.4). L'efficacité de ces estimateurs doit être comprise au sens local asymptotique, sous la structure de la normalité asymptotique locale (LAN) (cf. le chapitre 1). L'approche la plus commune pour l'efficacité est de consi-

dérer un sous-modèle paramétrique, dit sous-modèle le moins favorable, du modèle semi-paramétrique de départ (c'est un sous-modèle pour lequel l'information de Fisher est la plus petite) et d'estimer α^* dans un tel modèle (voir Van der Vaart (1998, 2002) pour les modèles semi-paramétrique identiquement distribués, et McNeney et Wellner (2000) pour les modèles semi-paramétriques non identiquement distribués). Ici les sous-modèles considérés sont des sous-modèles paramétriques dans lesquels la fonction f^* est un polynôme trigonométrique. Certains de nos processus de contraste sont définis à partir de la vraisemblance profilée dans ces sous-modèles (cf la section 2.2). La vraisemblance profilée est une méthode semi-paramétrique qui consiste à "éliminer" le composant non paramétrique f^* afin de ramener le problème semi-paramétrique original en un problème paramétrique convenablement choisi. Ces processus de contraste, construits à partir de la vraisemblance profilée, fournissent ainsi des estimateurs asymptotiquement efficaces. En cela, la procédure d'estimation est proche de la procédure d'estimation efficace de Gassiat et Lévy-Leduc (2006), par laquelle les auteurs estiment la période d'une fonction inconnue. De plus, ceci rejoint les travaux de Murphy et Van der Vaart (2000) dans les modèles semi-paramétriques indépendamment et identiquement distribués : la vraisemblance profilée sélectionne le modèle le moins favorable.

La vraisemblance profilée induit par définition un estimateur pour la forme commune f^* . Les propriétés asymptotiques de cet estimateur sont étudiés dans la sous-section 2.3.3 : pour $d = 1$ sa vitesse de convergence est la vitesse non paramétrique optimale (corollaire 2.1). Par contre, il n'est pas asymptotiquement efficace au sens de la proposition 1.2.

Le plan du chapitre est le suivant. La section 2.1 décrit le modèle étudié et les différentes contraintes d'identifiabilité considérées. La section 2.2 étudie la transformée de Fourier du modèle, et définit les processus de contraste et les estimateurs. Ensuite, la section 2.3 est consacrée à l'étude de l'efficacité asymptotique des estimateurs définis à partir de la vraisemblance profilée, et du comportement asymptotique de l'estimateur de la forme commune. Enfin, puisque chacun des processus de contraste est défini à partir d'un filtre linéaire sur les coefficients de Fourier, la section 2.4 discute du choix de ces filtres à partir de données simulées.

2.1 Présentation du modèle

Dans cette partie, on décrit le modèle étudié, ainsi que les contraintes d'identifiabilité considérées tout au long du chapitre.

2.1.1 Description du modèle

Les données $(Y_{i,j})_{i,j}$ sont issues de l'observation de J unités sur le domaine d'observation $\mathcal{D} = \prod_{r=1}^d [0, T_r]$. L'indice i représente l'indice des temps d'observation. L'indice j représente l'indice des unités (ou sujets). Quitte à faire un changement de variable on suppose que pour tout $r = 1, \dots, d$, $T_r = 2\pi$. Le modèle de régression étudié est alors le suivant :

$$Y_{i,j} = a_j^* f^*(t_{i,j} - \theta_j^*) + v_j^* + \sigma^* \epsilon_{i,j}, \quad i \in \mathcal{I}_{n,j}, \quad j = 1, \dots, J. \quad (2.3)$$

C'est un modèle de dépendance non linéaire, puisque généralement la fonction de régression f^* dépend non linéairement des points d'observation connus $t_{i,j} \in \mathcal{D}$. Cette fonction est supposée continue et périodique sur le réseau $2\pi\mathbb{Z}^d$, et on note \mathcal{F} l'ensemble des fonctions continues périodiques sur ce réseau. La fonction $f^* \in \mathcal{F}$ est appelée la forme commune du modèle. Elle est inconnue.

Les paramètres fini-dimensionnels $\theta^* = (\theta_{j,q}^*)_{1 \leq j \leq J, 1 \leq q \leq d}$, $a^* = (a_j^*)_{1 \leq j \leq J}$ et $v^* = (v_j^*)_{1 \leq j \leq J}$ caractérisent les unités j , $j = 1 \dots J$. Ils sont également inconnus. Au vu des données, il est raisonnable de supposer que ces paramètres appartiennent à un ensemble compact \mathcal{A} :

$$\alpha^* = (\theta^*, a^*, v^*) \in \mathcal{A} = [\theta_{\min}, \theta_{\max}]^{dJ} \times ([a_{\min}, a_{\max}] \setminus \{0\})^J \times [v_{\min}, v_{\max}]^J,$$

où $\theta_{\min} < \theta_{\max}$, $a_{\min} < a_{\max}$ et $0 < v_{\max} = -v_{\min}$ sont des réels fixés le plus souvent par le statisticien. Les paramètres $\theta_j^* = (\theta_{j,q}^*)_{1 \leq q \leq d} \in \mathbb{R}^d$ ($j = 1, \dots, J$) sont appelés paramètres de translation. Les paramètres a_j^* ($j = 1, \dots, J$) sont appelés paramètres d'amplitude. Les paramètres v_j^* ($j = 1, \dots, J$) sont appelés paramètres de niveau.

Les termes d'erreur $(\epsilon_{i,j})_{i,j}$ sont des variables indépendantes centrées de variance égale à 1. Dans la section 2.3 (où l'on considère l'estimation efficace de α^*), on suppose la loi des $(\epsilon_{i,j})_{i,j}$ connue : les termes d'erreur suivent une loi gaussienne.

Afin de simplifier les calculs, le modèle remplit les trois hypothèses présentées ci-dessous.

- Les points d'observation $t_{i,j}$ sont identiques pour chaque sujet j . On note :

$$t_{i,j} = t_i \in [0, 2\pi[^d, \quad i \in \mathcal{I}_n, \quad j = 1 \dots J, \quad (2.4)$$

avec $\mathcal{I}_n = \{1 \dots n\}^d$. On suppose également que ces points sont uniformément répartis sur le domaine d'observation et que :

$$t_i = \frac{i}{n} 2\pi, \quad i \in \mathcal{I}_n, \quad j = 1 \dots J. \quad (2.5)$$

Le fait de prendre des points équidistants permet de conserver l'orthogonalité de la base de Fourier (2.9) et par conséquent de simplifier les calculs. La remarque

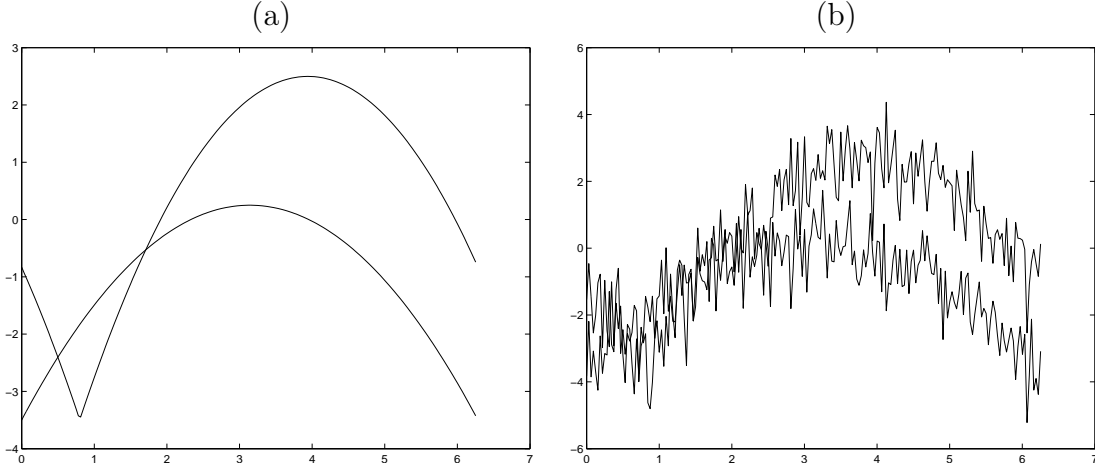


FIG. 2.1: Illustration du modèle (2.3) avec $d = 1$ et $J = 2$ courbes. Le graphique (a) représente deux fonctions vérifiant (2.2). La figure (b) représente ces mêmes fonctions bruitées avec $\sigma^* = 0.8$. La forme commune est générée par la fonction $f^*(x) = 20(1 - x/(2\pi))x/(2\pi)$.

2.2 explique brièvement le comportement des estimateurs que l'on définit dans le cas où les points d'observation t_i ne seraient pas équidistants.

- Le nombre d'observations par unité est impair (afin de garantir l'orthogonalité de la base de Fourier discrète (2.9)) :

$$(N1) \quad n \text{ est impair.}$$

- Les écarts-type σ_j^* , $j = 1 \dots J$ sont égaux :

$$\sigma_j^* = \sigma^*, \quad j = 1 \dots J.$$

Néanmoins une étude similaire peut être menée dans le cas où les écarts-type ne sont pas égaux (voir la Remarque 2.4).

Les figures 2.1(a) et 2.1(b) illustrent le modèle (2.3) pour $d = 1$. La figure 2.1(a) représente les fonctions que l'on souhaite observer. La première courbe est l'identique de la seconde à une translation, à une amplitude et à un niveau près. La figure 2.1(b) représente les données effectivement mesurées : un bruit blanc a été ajouté aux données théoriquement observées de la figure 2.1(a).

Le modèle étudié (2.3) est un sous-modèle du modèle de forme invariante défini par Lawton et al. (1972). C'est un modèle semi-paramétrique. Le paramètre fini-dimensionnel est le couple (α^*, σ^*) . Le paramètre infini-dimensionnel ou de nuisance est la fonction f^* . L'objectif de ce chapitre est de développer des méthodes d'estimation pour les paramètres α^* et f^* .

On estime le paramètre α^* en minimisant un processus de contraste (cf. section 1.1). Ces processus de contraste, comme le précise la sous-section 2.1.2, dépendent des contraintes d'identifiabilité posées sur le modèle. Quelles que soient ces contraintes, ils permettent de définir des estimateurs asymptotiquement normaux, indépendamment de la distribution des termes d'erreur $(\epsilon_{ij})_{i,j}$ (théorème 2.2). De plus, si la distribution des termes d'erreur est gaussienne, certains processus fournissent des estimateurs asymptotiquement efficaces (théorème 2.4).

2.1.2 Identifiabilité du modèle

Afin de pouvoir estimer les paramètres α^* et f^* , il faut que le modèle soit identifiable. Or ce n'est pas le cas. En effet, soient $\alpha = (\theta, a, v) \in \mathcal{A}$ et $f \in \mathcal{F}$ tels que :

$$\theta = \theta^* + \theta_0 + 2\pi k, \quad a = \lambda a^*, \quad v = v^* + \lambda a^* v_0, \quad f(\cdot) = \lambda^{-1} f^*(\cdot + \theta_0) - v_0, \quad (2.6)$$

où $\theta_0 \in \mathbb{R}^{dJ}$, $k \in \mathbb{Z}^{dJ}$, $\lambda \in \mathbb{R}^*$ et $v_0 \in \mathbb{R}$. Alors, pour tout $j = 1, \dots, J$, on a,

$$a_j f(\cdot - \theta_j) + v_j = a_j^* f^*(\cdot - \theta_j^*) + v_j^*.$$

Il convient donc de poser des contraintes d'identifiabilité sur ce modèle. Ces contraintes vont restreindre l'espace des paramètres \mathcal{A} et \mathcal{F} de telle sorte que :

- α^*, f^* vérifient ces contraintes ;
- si α, f vérifient ces contraintes et l'équation (2.6), alors $\alpha = \alpha^*$ et $f = f^*$;
- la fonction de contraste associée au paramètre α admet un unique minimum en α^* .

Dans ce chapitre, on considère trois jeux de contraintes d'identifiabilité :

jeu n°1 : $f \in \mathcal{F}$ et $\alpha \in \mathcal{A}_1 = \{\alpha \in \mathcal{A}, \text{ tel que } \theta_1 = 0, a_1 = 1 \text{ et } v_1 = 0\}$, où $\theta_{min} = 0$, $\theta_{max} = 2\pi$, $0 < a_{min} < 1 < a_{max}$,

Afin de simplifier les notations on suppose que pour les jeux de contraintes n°2 et n°3, on pose : $\theta_{min} = 0$, $\theta_{max} = 2\pi$, $a_{min} = -\sqrt{J}$, et $a_{max} = \sqrt{J}$.

jeu n°2 : $f \in \mathcal{F}$ et $\alpha \in \mathcal{A}_2 = \{\alpha \in \mathcal{A}, \text{ tel que } \theta_1 = 0, v_1 = 0, \text{ et } \sum_{j=1}^J a_j^2 = J, a_1 > 0\}$,

jeu n°3 : $f \in \mathcal{F}_0$ et $\alpha \in \mathcal{A}_3 = \{\alpha \in \mathcal{A}, \text{ tel que } \theta_1 = 0 \text{ et } \sum_{j=1}^J a_j^2 = J, a_1 > 0\}$,

où \mathcal{F}_0 désigne l'ensemble des fonctions continues périodiques telles que :

$$\int_{\mathcal{D}} f(t) \frac{dt}{(2\pi)^d} = 0.$$

Le modèle associé au jeu n°1 de contraintes signifie que la première unité est prise comme référence, c'est-à-dire que les unités $j = 2, \dots, J$ sont des versions de l'unité 1 à une translation, à une amplitude et un niveau près. Ce modèle est étudié dans la section 2.2.2 pour $d = 1$.

Les modèles associés au jeu n°2 et au jeu n°3 de contraintes sont ceux obtenus en utilisant la vraisemblance profilée du modèle. Dans ces deux modèles, l'unité $j = 1$ est prise comme référence pour estimer la translation. La condition d'identifiabilité sur le paramètre d'amplitude ($\sum_{j=1}^J a_j^2 = J$) (ici la constante J est choisie pour simplifier les calculs) s'impose si l'on considère la vraisemblance du modèle (voir la remarque 2.5). Elle signifie que la forme commune est définie comme la moyenne pondérée des fonctions de régression f_j^* . Cette condition d'identifiabilité permet de définir des estimateurs efficaces pour le paramètre α^* (cf le théorème 2.4).

Ce qui différencie les jeux de contraintes n°2 et n°3 est la condition posée sur la forme commune et le paramètre de niveau. Pour le jeu n°3, aucune contrainte n'est posée sur le niveau, par contre la forme commune appartient à l'espace de fonctions \mathcal{F}_0 . Au contraire, pour le jeu n°2, c'est sur le paramètre de niveau qu'est posée la contrainte et non sur la forme commune. Comme le souligne la remarque 2.6, le jeu de contrainte n°3 est le plus optimal en ce sens qu'il fournit non seulement des estimateurs asymptotiquement efficaces mais aussi asymptotiquement indépendants pour les paramètres θ^* , a^* et v^* . C'est pourquoi nous développons uniquement les calculs pour le modèle associé à ce dernier jeu (les calculs étant similaires) dans la section 2.2.3 et la section 2.3.

2.2 Définition des processus de contraste

2.2.1 Transformée de Fourier discrète du modèle

Puisque la transformée de Fourier sur le tore $\mathbb{T}^d = (\mathbb{R}/(2\pi\mathbb{Z}))^d$ est linéaire, et puisque la base de Fourier diagonalise l'opérateur de translation, cette transformée est bien adaptée au modèle étudié. En effet, si les fonctions de régression sont intégrables et vérifient l'équation (2.2), alors les coefficients de Fourier vérifient :

$$c_l(f_j^*) = \int_{\mathcal{D}} f_j^*(t) e^{-il \cdot t} \frac{dt}{(2\pi)^d} = \begin{cases} a_j^* c_0(f^*) + v_j^* & \text{si } l = 0, \\ a_j^* c_l(f^*) e^{-il \cdot \theta_j^*} & \text{si } l \in \mathbb{Z}^d \setminus \{0\}. \end{cases} \quad (2.7)$$

Nous développons ici l'estimation des coefficients de Fourier des fonctions f_j^* à partir des données $(Y_{i,j})$. Ces estimateurs sont utilisés tout au long du chapitre pour estimer les paramètres θ^* , a^* et v^* .

Une approche naturelle pour estimer une intégrale sur un compact est d'utiliser les formules de quadrature associées au design des points d'observation (see [Haber, 1970](#)) :

$$\int_{\mathcal{D}} f(t) \frac{dt}{(2\pi)^d} \simeq \sum_{i \in \mathcal{I}_n} \lambda_i f(t_i), \quad (2.8)$$

où λ_i sont des réels qui dépendent du choix des points d'observation (t_i) . Dans le cas où les points d'observation sont équidistants (2.5), cette formule est celle de Newton-Cotes avec $\lambda_i = 1/n^d$. L'approximation des coefficients de Fourier par cette dernière formule de quadrature est connue sous le nom de transformée de Fourier discrète (ou DFT comme "Discrete Fourier Transform") :

$$c_l^n(f) = \frac{1}{n^d} \sum_{s \in \mathcal{I}_n} f(t_s) e^{-il \cdot t_s}, \quad |l|_\infty < \frac{n}{2},$$

où $|l|_\infty = \max_{r=1 \dots d} |l_r|$ pour $l = (l_1, \dots, l_d) \in \mathbb{Z}^d$.

Le fait de considérer que les points t_i sont équidistants permet de conserver pour la transformée de Fourier discrète l'orthogonalité de la base de Fourier :

$$\forall |l|_\infty < \frac{n}{2}, \quad \forall |p|_\infty < \frac{n}{2}, \quad \frac{1}{n^d} \sum_{r \in \mathcal{I}_n} e^{i(l-p) \cdot t_r} = \begin{cases} 1 & \text{si } l = p, \\ 0 & \text{si } l \neq p. \end{cases} \quad (2.9)$$

Par conséquent, on peut établir la qualité de l'approximation de la transformée de Fourier par la DFT. Soit f une fonction continue telle que sa série de Fourier converge normalement. L'équation (2.9) permet d'établir un lien entre la DFT et la série de Fourier,

$$c_l^n(f) = c_l(f) + \sum_{p-l \in (n\mathbb{Z})^d \setminus \{0\}} c_p(f), \quad \forall |l|_\infty < \frac{n}{2}. \quad (2.10)$$

Toutefois si les points d'observation ne sont pas équidistants, la base de Fourier discrète n'est plus nécessairement orthogonale, mais elle le reste bien évidemment asymptotiquement (voir la remarque 2.2).

Un estimateur des coefficients de Fourier est alors :

$$d_{j,l} = \frac{1}{n^d} \sum_{s \in \mathcal{I}_n} e^{-il \cdot t_s} Y_{s,j} = c_l^n(f_j^*) + w_{j,l}, \quad j = 1 \dots J, \quad |l|_\infty < n/2. \quad (2.11)$$

Du fait de la propriété d'orthogonalité (2.9), les termes d'erreur $(w_{j,l})_{j,l}$ sont des variables complexes décorréées de variance $1/n^d$. Par ailleurs, pour j fixé et lorsque $l \neq 0$ ($w_{j,0}$ est une variable réelle), les parties réelle et imaginaire de $w_{j,l}$ sont décorréées et de variance $1/(2n^d)$.

Par conséquent, si les variables (ϵ_{ij}) sont gaussiennes, alors les variables $(\xi_{j,l} = n^{d/2} w_{j,l})$ sont indépendantes identiquement distribuées de loi $\mathcal{N}^{\mathbb{C}}(0, 1)$ (cf. l'annexe

A), c'est-à-dire que les parties réelle et imaginaire de $\xi_{j,l}$ sont indépendantes de loi gaussienne $\mathcal{N}(0, 1/2)$.

Remarque 2.1. Si les $(\epsilon_{i,j})$ ne suivent pas une gaussienne, le lemme 2.1 établit que la famille $(\xi_{j,l} = n^{d/2} w_{j,l})$ de variables possèdent asymptotiquement les mêmes propriétés énoncées ci dessus pour le cas où les variables sont gaussiennes.

2.2.2 Définition empirique d'un processus de contraste.

Dans cette sous-section, on présente un processus de contraste pour estimer le paramètre α^* . Ce processus est défini à partir de la relation (2.7). Afin d'alléger les calculs et les notations, on restreint l'étude pour $d = 1$ et pour le sous-modèle de (2.3) suivant :

$$Y_{i,j} = a_j^* f^*(t_i - \theta_j^*) + \sigma^* \epsilon_{i,j}, \quad j = 1 \dots J, \quad i = 1 \dots n, \quad (2.12)$$

muni des contraintes d'identifiabilité du jeu n°1,

$$f \in \mathcal{F} \text{ et } (\theta, a) \in \mathcal{A}_1 = \{(\theta, a) \in [0, 2\pi[\times [a_{\min}, a_{\max}[, \text{ tel que } \theta_1 = 0, \quad a_1 = 1\},$$

avec $0 < a_{\min} < 1 < a_{\max}$.

Si la série de Fourier de f^* est normalement convergente, l'équation (2.10) permet d'établir, uniformément pour $|l| < n/2$, la convergence des coefficients de Fourier discret ($c_l^n(f^*)$) vers la série de Fourier de f^* . D'autre part, l'introduction d'une suite de poids dans la définition du critère (2.16) permet de contrôler le biais entre les coefficients de Fourier discrets et les vrais coefficients de Fourier de f^* . On peut alors considérer la version simplifiée suivante de la DFT du modèle où les coefficients $c_l^n(f_j^*)$ sont remplacés par leur version continue :

$$d_{j,l} = a_j^* e^{-il\theta_j^*} c_l(f^*) + w_{j,l}, \quad j = 1 \dots J, \quad l = -(n-1)/2 \dots (n-1)/2. \quad (2.13)$$

Les théorèmes (2.1) et (2.2) sont établis pour le modèle simplifié (2.13). Néanmoins, ils sont également vrais pour le modèle original (2.3) si la série de Fourier de f^* est normalement convergente. Notons que dans la section 2.2.3 nous étudions le modèle 2.11) et non la version simplifiée considérée ici.

Définition du processus de contraste

Au vu de l'équation (2.7), l'idée est d'estimer les coefficients de Fourier de la forme commune f^* en normalisant le DFT des données $(d_{j,l})$, puis de les comparer à leurs moyennes. On définit les coefficients de Fourier normalisés,

$$\tilde{c}_{j,l}(\theta, a) = d_{j,l} e^{il\theta_j} / a_j, \quad j = 1, \dots, J, \quad |l| < n/2, \quad (2.14)$$

et leurs moyennes,

$$\hat{c}_l(\theta, a) = \frac{1}{J} \sum_{j=1}^J \tilde{c}_{j,l}(\alpha), \quad |l| < n/2.$$

Ainsi quand $(\theta, a) = (\theta^*, a^*)$, les coefficients $c_{j,l}(\theta^*, a^*)$ et $\hat{c}_l(\theta^*, a^*)$ sont tous deux proches de $c_l(f^*)$. Par conséquent, $|\tilde{c}_{j,l}(\alpha) - \hat{c}_l(\alpha)|^2$ devrait être petit quand (θ, a) est dans un voisinage proche de (θ^*, a^*) . Le critère d'estimation est défini à partir de ces différences et d'une suite de réels positifs $(\delta_l)_{l \in \mathbb{Z}}$ de carré sommable qui pondèrent les coefficients de Fourier,

$$\sum_{l \in \mathbb{Z}} \delta_l^2 < +\infty \quad \text{et} \quad \delta_l = \delta_{-l} \quad \forall l \in \mathbb{Z}. \quad (2.15)$$

Le processus de contraste pour l'estimation de θ^* et a^* est alors le suivant,

$$W_n(\theta, a) = \frac{1}{J} \sum_{j=1}^J \sum_{l=-(n-1)/2}^{(n-1)/2} \delta_l^2 |\tilde{c}_{j,l}(\theta, a) - \hat{c}_l(\theta, a)|^2. \quad (2.16)$$

Ce processus est clairement une fonction positive en (θ, a) , et il est construit de telle sorte que sa valeur minimale soit atteinte dans un voisinage proche de (θ^*, a^*) .

L'introduction d'une suite de poids permet ici de contrôler la partie stochastique du critère et le biais provenant de la considération du modèle simplifié (2.13) à la place du vrai modèle. Dans cette sous-section, la suite de poids est un filtre linéaire fixé (il n'évolue pas avec la taille des observations n) sur les coefficients de Fourier des fonctions f_j^* . On effectue sur les données la projection avec poids suivante :

$$(c_l)_l \in l^2(\mathbb{Z}) \longrightarrow (\delta_l c_l) \in l^2(\mathbb{Z}).$$

Cette projection se traduit par un opérateur de convolution sur l'espace des fonctions :

$$f \in \mathbb{L}^2(\mathcal{D}) \longrightarrow \psi * f \in \mathbb{L}^2(\mathcal{D}),$$

où ψ est la \mathbb{L}^2 -fonction à valeur réelle définie par :

$$\psi(t) = \sum_{l \in \mathbb{Z}} \delta_l e^{ilt}. \quad (2.17)$$

Le choix le plus naturel est de prendre une suite (δ_l) ayant un nombre fini de coefficients non nuls, comme :

$$\delta_l = \begin{cases} 1 & \text{si } |l| \leq m \\ 0 & \text{si } |l| > m, \end{cases}$$

puis d'étudier le critère quand l'entier m croît avec n . Une telle étude est menée dans la sous-section suivante et se poursuit dans la section 2.3. Intuitivement, un choix optimal serait de considérer une suite de poids dont les coefficients seraient nuls pour les coefficients de Fourier de f^* négligeables. Ce choix dépendrait donc de la forme commune qui est inconnue. Des techniques ont été développées pour choisir le poids le plus optimal (voir par exemple Pinsker (1980), Tsybakov (2004)). Plus récemment, Dalalyan et al. (2006) développent une approche bayésienne pour l'estimation du centre de symétrie d'une fonction périodique. De plus, pour ce dernier modèle, Dalalyan (2005) construit une suite de poids définie par bloc (méthode bayésienne de Stein). Les poids sont calculés à partir de la loi à priori sur les coefficients de Fourier de f^* . Ce type de méthodes a été aussi utilisé pour l'estimation de la période pour une fonction périodique par Castillo (2005).

Les résultats

Afin d'assurer l'unicité du minimum pour la fonction de contraste, on doit s'assurer que la fonction f^* est réellement 2π -périodique :

$$2\pi = \min\{T \in \mathbb{R}_+^*, \forall t \in \mathbb{R} f(t+T) = f(t)\}.$$

On est donc contraint à supposer que la suite de poids et la forme commune vérifient au moins l'une de ces deux hypothèses,

$$(W1) \quad \text{les coefficients } c_1(f^*) \text{ et } \delta_1 \text{ sont non nuls.}$$

ou

$$(W2) \quad \text{il existe } (l, p) \in \mathbb{Z}^2 \text{ des entiers premiers entre eux (le plus grand diviseur commun est 1) tels que } c_l(f^*), c_p(f^*), \delta_l \text{ et } \delta_p \text{ soient non nuls.}$$

Si on définit l'estimateur de (θ^*, a^*) comme le minimum du processus de contraste (2.16), on montre que cet estimateur est consistant.

Théorème 2.1. *Soit $(\hat{\theta}_n, \hat{a}_n)$ l'estimateur minimisant le processus de contraste (2.16) :*

$$(\hat{\theta}_n, \hat{a}_n) = \arg \min_{\beta \in \mathcal{A}_1} W_n(\beta).$$

Si la suite de poids (δ_l) et la forme commune f^ sont telles que l'une des hypothèses (W1) ou (W2) soit vérifiée, et que*

$$\sum_{l \in \mathbb{Z}} |c_l(f)|^2 < \infty \quad \text{et} \quad \sum_{l \in \mathbb{Z}} \delta_l^2 < \infty \tag{2.18}$$

$(\hat{\theta}_n, \hat{a}_n)$ converge en probabilité vers (θ^, a^*) .*

Démonstration. La démonstration de ce théorème revient à montrer les deux points suivants (cf le chapitre 1, section 1.1) :

- i) le processus de contraste converge (2.16) uniformément en probabilité vers une fonction de contraste K , (voir le lemme 2.2)

$$K(\theta, a) = \frac{1}{J} \sum_{j=1}^J |a_j^*/a_j|^2 \int_0^{2\pi} |\psi * f(t)|^2 \frac{dt}{2\pi} - \int_0^{2\pi} \left| \sum_{j=1}^J \frac{a_j^*}{a_j} \psi * f(t + \theta_j - \theta_j^*) \right|^2 \frac{dt}{2\pi};$$

- ii) K admet un unique minimum en (θ^*, a^*) (voir le lemme 2.3);

□

On établit également la \sqrt{n} -normalité asymptotique de cet estimateur sous les hypothèses décrites ci dessous.

Théorème 2.2. *Considérons l'estimateur de (θ^*, a^*) défini au théorème 2.1. Supposons que la suite de poids (δ_l) et la forme commune f^* sont telles que l'une des hypothèses (W1) ou (W2) soit vérifiée, et que*

$$\sum_{l \in \mathbb{Z}} |c_l(f)|^2 < \infty, \quad (2.19)$$

$$\sum_{l \in \mathbb{Z}} \delta_l^2 |l|^2 < \infty. \quad (2.20)$$

Alors la suite des fluctuations $\sqrt{n}(\hat{\theta}_n - \theta^*, \hat{a}_n - a^*)$ converge en distribution vers un vecteur Gaussien $\mathcal{N}_{2J-2}(0, \sigma^{*2}\Gamma)$, où Γ est la matrice suivante,

$$\Gamma = \frac{J^2}{2} \begin{pmatrix} \frac{\sum_{l \in \mathbb{Z}} \delta_l^4 l^2 |c_l(f)|^2}{(\sum_{l \in \mathbb{Z}} \delta_l^2 l^2 |c_l(f)|^2)^2} \mathbf{V} & 0 \\ 0 & \frac{\sum_{l \in \mathbb{Z}} \delta_l^4 |c_l(f)|^2}{(\sum_{l \in \mathbb{Z}} \delta_l^2 |c_l(f)|^2)^2} \mathbf{D}_a \mathbf{V} \mathbf{D}_a \end{pmatrix},$$

et $V = \frac{1}{J^2} \left(D_a^{-2} + \frac{1}{a_1^{*2}} \mathbb{I}_{J-1} (\mathbb{I}_{J-1})^T \right)$, $D_a = \text{diag}(a_2^* \dots a_J^*)$, et $\mathbb{I}_{J-1}^T = (1 \dots 1)^T \in \mathbb{R}^{J-1}$.

Démonstration. Par définition, la fonction W_n est deux fois différentiable sur l'espace des paramètres. Or le M -estimateur est défini comme le minimum du processus de contraste (2.16). Par conséquent, $(\hat{\theta}_n, \hat{a}_n)$ est solution du système $\nabla W_n(\hat{\alpha}_n) = 0$, où ∇ est l'opérateur gradient. Par un développement de Taylor du gradient, il existe $\bar{\alpha}_n$ dans un voisinage proche de α^* tel que,

$$\nabla^2 W_n(\bar{\alpha}_n) n^{1/2}(\hat{\alpha}_n - \alpha^*) = -n^{1/2} \nabla W_n(\alpha^*),$$

où ∇^2 est l'opérateur hessien. D'après les propositions 2.3 et 2.4, $n^{1/2} \nabla W_n(\theta^*, a^*)$ converge en loi vers un vecteur gaussien centré de matrice de covariance Γ_0 , et la

matrice $\nabla^2 M_n(\bar{\alpha}_n)$ converge en probabilité vers une matrice inversible H . On en déduit par le lemme de Slutsky (Van der Vaart (1998) 2.8) que les fluctuations de l'estimateur convergent vers un vecteur gaussien centré de matrice de covariance $\Gamma = H^{-1}\Gamma_0(H^{-1})^T$. \square

Notons que la suite de poids étant ici choisie arbitrairement, la régularité du critère est reportée sur celle-ci et non sur la forme commune. Si on considère la version simplifiée du modèle (2.13), l'hypothèse (2.19) équivaut à supposer que la forme commune soit de carré intégrable sur $[0, 2\pi]$. Pour le modèle (2.12), le théorème 2.2 est vrai si l'hypothèse (2.19) est remplacée par :

$$\sum_{l \in \mathbb{Z}} |c_l(f^*)| < \infty.$$

Cette dernière condition est légèrement plus forte que la simple continuité de f^* .

Remarque 2.2. *Pour le modèle où les temps d'observation ne sont pas équidistants, le choix de la formule de quadrature (2.8) influe sur la qualité de l'estimateur. Le théorème de consistance peut être établi si les coefficients λ_i satisfont les propriétés suivantes :*

$$\begin{aligned} \lim_{n \rightarrow \infty} \sup_{i=1 \dots n} \lambda_i &= 0, \\ \lim_{n \rightarrow \infty} \sup_{0 < |l| \leq m_n} \left| \sum_{r=1}^n \lambda_r^2 e^{ilt_r} / \left(\sum_{s=1}^n \lambda_s^2 \right) \right| &= 0, \end{aligned} \quad (2.21)$$

où la suite $(m_n)_n$ est une suite croissante d'entiers tendant vers l'infini avec n (elle dépend du choix des temps d'observation). L'hypothèse (2.21) garantit l'uniforme convergence de la base $(e^{ilt_i})_{i=1 \dots n}$, $|l| \leq m_n$ vers la base orthogonale de Fourier. La normalité asymptotique peut être établie également sauf que la vitesse de convergence de l'estimateur dépend de la qualité de la formule de quadrature. Elle est égale ici à $(\sum_{s=1}^n \lambda_s^2)^{-1}$.

2.2.3 Définition d'un processus de contraste à partir de la vraisemblance profilée

Le processus de contraste précédemment étudié est défini à partir de la relation (2.7) de façon empirique. Dans cette sous-section, les processus de contraste sont définis à partir de la vraisemblance profilée (cf. le chapitre 1). Le modèle étudié est celui qui est présenté dans la section 2.1 muni du jeu n°3 de contraintes d'identifiabilité (voir la sous-section 2.1.2) :

$$f \in \mathcal{F}_0 \quad \text{et} \quad \alpha = (\theta, a, v) \in \mathcal{A}_3.$$

On suppose dorénavant que les variables d'erreur $(\epsilon_{i,j})$ sont des variables gaussiennes centrées réduites.

Définition du processus de contraste

La log-vraisemblance l_n du modèle est alors égale à,

$$l_n(\alpha, f) = \frac{-1}{2\sigma^2} \sum_{i \in \mathcal{I}_n} \sum_{j=1}^J (Y_{i,j} - a_j f(t_i - \theta_j) - \nu_j)^2 - \frac{nJ}{2} \log \sigma^2. \quad (2.22)$$

Considérer la vraisemblance profilée consiste à maximiser la vraisemblance en deux fois. D'abord on maximise par rapport au paramètre de nuisance f , on calcule ainsi la vraisemblance profilée du modèle :

$$pl_n(\alpha, \sigma^2) = \sup_{f \in \mathcal{F}_0} l_n(\alpha, \sigma^2, f).$$

La seconde étape est de maximiser le processus de contraste pl_n pour estimer le paramètre (α, σ^2) . Remarquons que les estimateurs du maximum de vraisemblance pour α^* et σ^{*2} , les deux premières composantes du triplet $(\hat{\alpha}_n, \hat{\sigma}_n^2, \hat{f}_n)$ qui maximise $l_n(\alpha, \sigma^2, f)$, sont le maximum de la vraisemblance profilée $(\alpha, \sigma^2) \rightarrow pl_n(\alpha, \sigma^2)$.

Mais, maximiser la vraisemblance sur l'espace \mathcal{F}_0 des fonctions continues périodiques semble néanmoins assez compliqué. On réduit alors le problème semi-paramétriques en considérant une suite croissante de modèles paramétriques qui approche le modèle entier. Cette suite de modèles est ici la restriction du modèle (2.3) au cas où f est un polynôme trigonométrique dont le degré augmente avec le nombre d'observations n . Plus précisément, soit (m_n) une suite croissante d'entiers. On pose $\mathcal{F}_{0,n}$ le sous-espace de \mathcal{F}_0 des polynômes trigonométriques dont le degré partiel pour chacune des variables est inférieur à m_n . Pour toute fonction f de $\mathcal{F}_{0,n}$, il existe donc un vecteur complexe $(c_l)_{1 \leq |l| \leq m_n} \in \mathbb{C}^{(2m_n+1)^d-1}$ tel que :

$$c_l = \overline{c_{-l}} \quad \forall |l|_\infty \leq n/2, \quad \text{et} \quad f(t) = \sum_{1 \leq |l|_\infty \leq m_n} c_l e^{il \cdot t} \quad \forall t \in \mathbb{R}^d.$$

Afin de conserver l'orthogonalité de la base de Fourier (2.9), (m_n) et n vérifient les conditions suivantes :

$$(N2) \quad 2|m_n| < n, \quad \lim_{n \rightarrow +\infty} m_n = +\infty \quad (2.23)$$

Ainsi après calcul, le maximum de la vraisemblance sur l'espace $\mathcal{F}_{0,n}$ est atteint en la fonction \hat{f}_α ,

$$\hat{f}_\alpha(t) = \sum_{1 \leq |l|_\infty \leq m_n} \hat{c}_l(\alpha) e^{il \cdot t} \quad \forall t \in \mathbb{R}^d, \quad (2.24)$$

avec pour $l \in \mathbb{Z}^d$, $1 \leq |l|_\infty \leq m_n$,

$$\hat{c}_l(\alpha) = \left(n \sum_{j=1}^J a_j^2 \right)^{-1} \sum_{j=1}^J a_j \sum_{i=1}^n (Y_{ij} - v_j) e^{-il(t_i - \theta_j)}, \quad \forall \alpha \in \mathcal{A}_3. \quad (2.25)$$

On remplace la fonction f par \hat{f}_α dans la vraisemblance (2.22). Par la propriété d'orthogonalité de la base de la DFT (2.9), le premier terme de la vraisemblance profilée pl_n devient :

$$\begin{aligned} \sum_{j=1}^J \sum_{i \in \mathcal{I}_n} \left(Y_{ij} - a_j \sum_{1 \leq |l|_\infty \leq m_n} \hat{c}_l(\alpha) e^{il \cdot (t_i - \theta_j)} - v_j \right)^2 \\ = \sum_{j=1}^J \sum_{i \in \mathcal{I}_n} (Y_{ij} - v_j)^2 - \left(n \sum_{j=1}^J a_j^2 \right) \sum_{1 \leq |l|_\infty \leq m_n} |\hat{c}_l(\alpha)|^2. \end{aligned}$$

Posons L_n le processus de contraste pour l'estimation du paramètre α suivant :

$$L_n(\alpha) = \left(n \sum_{j=1}^J a_j^2 \right)^{-1} \sum_{j=1}^J \sum_{i \in \mathcal{I}_n} (Y_{ij} - v_j)^2 - \sum_{1 \leq |l|_\infty \leq m_n} |\hat{c}_l(\alpha)|^2. \quad (2.26)$$

La vraisemblance profilée se récrit alors de la façon suivante,

$$pl_n(\alpha, \sigma^2) = - \left(n \sum_{j=1}^J a_j^2 \right) \frac{L_n(\alpha)}{2\sigma^2} - \frac{nJ}{2} \log \sigma^2. \quad (2.27)$$

Remarquons que le processus de contraste L_n est par définition le critère des moindres carrés profilé du modèle.

Consistance des estimateurs

Afin d'assurer la consistance des résultats, on doit s'assurer que la fonction f^* est bien une fonction périodique sur le réseau $2\pi\mathbb{Z}^d$. Plus précisément, il ne doit pas exister une fonction $g : \mathbb{R}^{d-1} \rightarrow \mathbb{R}$ périodique sur $2\pi\mathbb{Z}^{d-1}$ telle que, par un changement de variable $\psi : t \in \mathbb{R}^d \rightarrow (z, x) \in \mathbb{R}^{d-1} \times \mathbb{R}$,

$$\forall t \in \mathbb{R}^d \quad g(z) = f(\psi(t)).$$

Par exemple pour $d = 2$, la fonction $f(t_1, t_2) = \cos(2(t_1 + t_2)) + \cos(3(t_1 + t_2))$ ne vérifie pas cette hypothèse. Il faut aussi que 2π soit la période minimale de la fonction, c'est-à-dire :

$$\forall r = 1, \dots, d, \quad 2\pi = \min\{T, \forall (t_1, \dots, t_d) \in \mathbb{R}^d \quad f(t_1, \dots, t_r + T, \dots, t_d) = f(t_1, \dots, t_r, \dots, t_d)\}.$$

Par exemple pour $d = 2$, la fonction $f(t_1, t_2) = \cos(2(t_1 + t_2)) + \cos(2t_1 + 3t_2)$ ne vérifie pas cette hypothèse pour $r = 1$. Ces propriétés sur f^* se traduisent sur les coefficients de Fourier par la vérification d'au moins une de ces deux hypothèses (elles sont les analogues des hypothèses **(W1)** et **(W2)** de la sous-section 2.2.2)

(F1) il existe une famille de d vecteurs de \mathbb{Z}^d , notée $\{l_r = (l_{1r}, \dots, l_{dr}) \in \mathbb{Z}^d, r = 1, \dots, d\}$, telle que $c_{l_r}(f^*)$ soit non nul pour tout $r = 1, \dots, d$ et que le déterminant de la matrice $(l_{sr})_{s,r}$ soit égal à ± 1 ,

ou

(F2) il existe deux familles de d vecteurs de \mathbb{Z}^d , $\{l_r = (l_{1r}, \dots, l_{dr}) \in \mathbb{Z}^d, r = 1, \dots, d\}$, et $\{p_r = (p_{1r}, \dots, p_{dr}) \in \mathbb{Z}^d, r = 1, \dots, d\}$, telles que $c_{l_r}(f^*)$ et $c_{p_r}(f^*)$ soient non nuls pour tout $r = 1, \dots, d$ et que les déterminants des matrices $(l_{sr})_{s,r}$ et $(p_{sr})_{s,r}$ soient premiers entre eux.

Sous ces conditions, les processus de contraste L_n et pl_n fournissent des estimateurs consistants. Il est à noter que pour la consistance de ces estimateurs, il n'est pas nécessaire de supposer que le bruit ϵ_{ij} est gaussien.

Théorème 2.3 (Consistance). *On considère les estimateurs définis par :*

$$\hat{\alpha}_n = \arg \min_{\alpha \in \mathcal{A}_3} L_n(\beta) \quad \text{et} \quad (\hat{\alpha}_n, \hat{\sigma}_n^2) = \arg \max_{(\alpha, \sigma^2) \in \mathcal{A}_3 \times \mathbb{R}_+^*} pl_n(\alpha, \sigma^2).$$

Supposons que la fonction f^* et la suite (m_n) vérifient les propriétés **(F1)** ou **(F2)**, **(N1)**, **(N2)** et :

$$\sum_{|l|_\infty > m} |c_l(f^*)| = o\left(\frac{1}{m^{d/2}}\right), \quad (2.28)$$

$$\frac{m_n^d}{n^d} = o(1), \quad (2.29)$$

alors $\hat{\alpha}_n$ et $\hat{\sigma}_n^2$ convergent en probabilité vers α^* .

Pour $d = 1$, l'hypothèse (2.28) est équivalente à supposer que la forme commune est une fonction hölderienne de régularité $1/2$. Pour $d \geq 2$, cette hypothèse signifie que la forme soit au moins $[d/2] + 1$ différentiable (où $[x] \in \mathbb{Z}$ désigne la partie entière du réel $x \in \mathbb{R}$). On peut néanmoins baisser cette exigence sur la régularité en réduisant le nombre η_n de coefficients de Fourier considéré dans le critère (dans le théorème ci dessus η_n est égale à $(2m_n)^d$). Par exemple si f^* est le produit de d fonctions f_1, \dots, f_d hölderiennes de régularité $\alpha < 1$, définies sur \mathbb{R} telle que pour tout $t = (t_1, \dots, t_d) \in \mathbb{R}^d$, $f^*(t) = f_1(t_1) \dots f_d(t_d)$, alors la consistance des estimateurs est établie si $\eta_n/n^{2\alpha}$ converge vers zéro.

Démonstration. Par le théorème de convergence des M -estimateurs (cf. le chapitre 1, section 1.1), il suffit de montrer que :

- i) la convergence uniforme de L_n vers la fonction de contraste $L = M + \sigma^{*2}$ (voir le lemme 2.6), où

$$M(\alpha) = \int_{\mathcal{D}} \frac{1}{J} \sum_{j=1}^J (f_j^*(t) - v_j)^2 \frac{dt}{2\pi} - \int_{\mathcal{D}} \left(\sum_{j=1}^J a_j a_j^* f_j^*(t - \theta_j^* + \theta_j) \right)^2 \frac{dt}{2\pi}.$$

- ii) $M(\cdot)$ admet un unique minimum en α^* (voir le lemme 2.7).

□

Remarque 2.3. Une étude similaire peut être menée pour le modèle associé au jeu n°2 de contraintes d'identifiabilité (voir la sous-section 2.1.2). Le processus de contraste associé à ce modèle est alors

$$L_n(\alpha) = \left(n \sum_{j=1}^J a_j^2 \right)^{-1} \sum_{j=1}^J \sum_{i \in \mathcal{I}_n} (Y_{ij} - v_j)^2 - \sum_{|l|_{\infty} \leq m_n} |\hat{c}_l(\alpha)|^2. \quad (2.30)$$

La différence avec le processus (2.26) est que le coefficient de Fourier pour $l = (0, \dots, 0) \in \mathbb{Z}^d$ intervient dans la définition. Ce critère converge uniformément vers la fonction de contraste suivante,

$$\int_{\mathcal{D}} \frac{1}{J} \sum_{j=1}^J (f_j^*(t) - v_j)^2 \frac{dt}{2\pi} - \int_{\mathcal{D}} \left(\sum_{j=1}^J a_j (f_j^*(t + \theta_j) - v_j) \right)^2 \frac{dt}{2\pi}.$$

Sous les mêmes hypothèses que le théorème 2.3, l'estimateur qui minimise le processus (2.30) est consistant. A priori, la fonction de contraste semble ici plus naturelle qu'avec le jeu de contraintes d'identifiabilité n°3 : elle est définie à partir des fonctions de régression f_j^* . Par contre, ce critère définit de moins bons estimateurs comme le souligne la remarque 2.6.

Remarque 2.4. Concernant le modèle où les écarts-type σ_j^* ne sont pas tous égaux, le processus de contraste (2.26) permet de définir un estimateur consistant pour le paramètre α^* . Mais cet estimateur ne s'avère pas asymptotiquement efficace par la suite. Pour construire un estimateur asymptotiquement efficace, on peut considérer

l'estimateur maximisant la log-vraisemblance profilée du modèle,

$$\begin{aligned}
 pl_n(\alpha, \sigma) = & - \sum_{i \in \mathcal{I}_n} \sum_{j=1}^J \frac{(Y_{ij} - v_j)^2}{\sigma_j^2} + \sum_{j=1}^J n^d \log \sigma_j^2, \\
 & + \sum_{1 \leq |l| \leq m_n} \sum_{j=1}^J n^d \left\{ 2\Re \left[\frac{\hat{c}_l(\alpha) \overline{\tilde{c}_{jl}(\alpha)}}{\sigma_j^2} \right] + \frac{a_j^2}{\sigma_j^2} |\hat{c}_l(\alpha)|^2 \right\} \\
 \text{où } \tilde{c}_{jl}(\alpha) = & \sum_{s=1}^n a_j (Y_{sj} - v_j) e^{-il(t_s - \theta_j)} / n^d, \quad \text{pour } 1 \leq |l| \leq m_n.
 \end{aligned}$$

Sous les mêmes hypothèses que le théorème précédent cet estimateur est consistant.

Remarque 2.5. La condition d'identifiabilité sur le paramètre a est essentielle pour que la fonction de contraste M admette un unique minimum en α^* . En effet si on reprend la preuve du lemme 2.7, la fonction M admet un unique minimum si, et seulement si les fonctions,

$$(\theta, a) \rightarrow \left(\sum_{j=1}^J a_j^2 \right)^{-2} \left| \sum_{j=1}^J a_j a_j^* e^{il \cdot (\theta_j - \theta_j^*)} \right|^2 \quad l \in \mathbb{Z}^*,$$

admettent un unique maximum (θ^*, a^*) (modulo 2π pour θ). La contrainte $(\sum_{j=1}^J a_j^2 = J)$ garantit cette propriété.

2.3 Efficacité et normalité asymptotique

Le but de cette section est d'étudier l'efficacité des estimateurs définis dans la section précédente. Le plan de l'étude est celui présenté au chapitre 1 : on montre d'abord que le modèle est LAN (voir la section 2.3.1), puis on étudie les propriétés asymptotiques des estimateurs définis à partir de la vraisemblance profilée.

2.3.1 Normalité Asymptotique Locale (LAN) du modèle

On considère dans cette sous-section le modèle (2.3) muni du jeu n°3 de contraintes d'identifiabilité (voir la sous-section 2.1.2). On suppose également que les variables d'erreurs (ϵ_{ij}) sont des variables gaussiennes indépendantes centrées réduites. Le modèle étudié est semi-paramétrique. Le paramètre fini-dimensionnel est le paramètre $(\alpha^*, \sigma^*) \in \mathcal{A}_3 \times \mathbb{R}_+$. Le paramètre de nuisance est la forme commune $f^* \in \mathcal{F}_0$.

Soient $(\alpha, \sigma) \in \mathcal{A}_3 \times \mathbb{R}_+$ et $f \in \mathcal{F}_0$. Pour tout $t \in \mathcal{D}$, on note $\mathbb{P}_{\alpha, \sigma, f}(t)$ le vecteur gaussien de \mathbb{R}^J de variance $\sigma^2 I_J$ et d'espérance le vecteur $(a_j f(t - \theta_j) + v_j)_{j=1 \dots J}$. Le

modèle théorique étudié est la collection de processus \mathcal{P} suivante :

$$\mathcal{P} = \{(\mathbb{P}_{\alpha,\sigma,f}(t))_{t \in \mathcal{D}}, (\alpha, \sigma) \in \mathcal{A}_3 \times \mathbb{R}_+, f \in \mathcal{F}_0\}.$$

On observe le modèle en un nombre fini de temps $t \in \mathcal{D}$. Pour $i \in \mathcal{I}_n$, notons $X_{i,n} = (Y_{i,1}, \dots, Y_{i,J})$, alors la distribution du vecteur $X_{i,n}$ est $\mathbb{P}_{\alpha,\sigma,f}(t_i)$. Ainsi, le modèle des observations est :

$$\mathcal{P}^{(n)} = \left\{ \mathbb{P}_{\alpha,\sigma,f}^{(n)} = \bigotimes_{i \in \mathcal{I}_n} \mathbb{P}_{(\alpha,\sigma,f)}(t_i) : (\alpha, \sigma, f) \in \mathcal{A}_3 \times \mathbb{R}_+ \times \mathcal{F} \right\}.$$

Avant d'étudier l'estimation efficace des paramètres, le modèle doit satisfaire des conditions de régularité : il doit être localement asymptotiquement normal (LAN). La propriété (LAN) requiert que le log-rapport de vraisemblance entre deux points $(\alpha^*, \sigma^*, f^*)$ et $(\alpha_n(h), \sigma_n(h), f_n(h))$, qui appartient à un voisinage $(\alpha^*, \sigma^*, f^*)$ dans la direction h , converge en loi vers une variable gaussienne qui dépend seulement de la direction h . La propriété (LAN) permet l'identification de la direction la moins favorable pour approcher le modèle, et donc de caractériser les estimateurs asymptotiquement efficaces. Comme les variables $X_{1,n}, \dots, X_{n,n}$ ne sont pas identiquement distribuées, l'étude présentée suit l'analyse semi-paramétrique développée par [McNeney et Wellner \(2000\)](#).

Soit $(\alpha, \sigma, f) \in \mathcal{A}_3 \times \mathbb{R}_+ \times \mathcal{F}$. On note le log-rapport de vraisemblance entre les deux points $(\alpha^*, \sigma^*, f^*)$ et (α, σ, f) par

$$\Lambda_n(\alpha, \sigma, f) = \log \frac{d\mathbb{P}_{\alpha,\sigma,f}^{(n)}}{d\mathbb{P}_{\alpha^*,\sigma^*,f^*}^{(n)}}.$$

La proposition suivante établit la propriété LAN du modèle. Notons $\mathbb{L}^2(\mathcal{D})$ l'espace de Hilbert des fonctions de carré intégrable sur \mathcal{D} .

Proposition 2.1. *Supposons que f^* soit continue non nulle et différentiable telle que ses dérivées partielles soient continues et forment une famille libre de l'espace vectoriel $\mathbb{L}^2(\mathcal{D})$. On note $\nabla f : \mathcal{D} \rightarrow \mathbb{R}^d$ la différentielle de f . Supposons que pour tout $j = 1 \dots J$, a_j^* soit non nul. Soit \mathcal{H} l'espace vectoriel suivant :*

$$\mathcal{H} = (\mathbb{R}^{d(J-1)} \times \mathbb{R}^{J-1} \times \mathbb{R}^J) \times \mathbb{R} \times \mathcal{F}_0,$$

muni du produit scalaire suivant $\langle \cdot, \cdot \rangle_{\mathcal{H}}$,

$$\forall h = (\alpha, \sigma, f) \in \mathcal{H}, \quad h' = (\alpha', \sigma', f') \in \mathcal{H},$$

$$\begin{aligned} \langle h, h' \rangle = & 2J \frac{\sigma \sigma'}{\sigma^{*2}} + \frac{1}{\sigma^{*2}} \sum_{j=1}^J v_j v'_j \\ & + \frac{1}{\sigma^{*2}} \left\langle a_1^* f - \sum_{j=2}^J a_j \frac{a_j^*}{a_1^*} f^*, a_1^* f' - \sum_{j=2}^J a'_j \frac{a_j^*}{a_1^*} f^* \right\rangle_{\mathbb{L}^2(\mathcal{D})} \\ & + \frac{1}{\sigma^{*2}} \sum_{j=2}^J \langle a_j^* f + a_j f^* - a_j^* \nabla f^* \cdot \theta_j, a_j^* f' + a'_j f^* - a_j^* \nabla f^* \cdot \theta'_j \rangle_{\mathbb{L}^2(\mathcal{D})}, \end{aligned}$$

où $\langle \cdot, \cdot \rangle_{\mathbb{L}^2(\mathcal{D})}$ est le produit scalaire de $\mathbb{L}^2(\mathcal{D})$. Alors le modèle (2.3) est LAN au point $(\alpha^*, \sigma^*, f^*)$ indexé par l'espace tangent \mathcal{H} . Plus précisément, pour tout $h \in \mathcal{H}$, il existe une suite $(\alpha_n(h), \sigma_n(h), f_n(h))$ telle que

$$\Lambda_n(\alpha_n(h), \sigma_n(h), f_n(h)) = \Delta_n(h) - \frac{1}{2} \|h\|_{\mathcal{H}}^2 + o_{\mathbb{P}_{\alpha^*, \sigma^*, f^*}}(1).$$

Ici le processus centré $\Delta_n(h)$ est linéaire avec h ,

$$\frac{1}{\sqrt{n}} \sum_{i \in \mathcal{I}_n} \sum_{j=1}^J \{(\sigma/\sigma^*)(\epsilon_{ij}^2 - 1) + A_{ij}^n(h) \epsilon_{ij}/\sigma^*\}, \quad (2.31)$$

où pour tout $i \in \mathcal{I}_n$,

$$A_{ij}^n(h) = \begin{cases} a_1^* f(t_i) - \sum_{k=2}^J a_k \frac{a_k^*}{a_1^*} f^*(t_i) & \text{si } j = 1 \\ a_j^* f(t_i - \theta_j^*) + a_j f^*(t_i - \theta_j^*) - a_j^* \nabla f^*(t_i - \theta_j^*) \cdot \theta_j + \nu_j & \text{si } j \in \{2 \dots J\}. \end{cases}$$

La preuve de cette proposition se trouve à la section 2.6.

Dans les modèles semi-paramétriques indépendants identiquement distribués, le fait que l'espace tangent \mathcal{H} ne soit pas complet n'implique pas l'existence d'une direction la moins favorable. Ainsi pour identifier cette direction, il est important que \mathcal{H} soit un sous-espace d'un espace de Hilbert muni du produit scalaire $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ (afin d'utiliser par la suite le théorème de représentation de Riesz). Ici, l'espace \mathcal{H} admet un complété pour le produit scalaire (voir le lemme 2.9) que l'on note $\text{cl}(\mathcal{H})$,

$$\text{cl}(\mathcal{H}) = (\mathbb{R}^{(J-1)d} \times \mathbb{R}^{J-1} \times \mathbb{R}^J) \times \mathbb{R} \times \{f \in \mathbb{L}^2(\mathcal{D}), c_0(f) = 0\}.$$

2.3.2 Estimation efficace du paramètre α^* .

On souhaite ici estimer efficacement le paramètre ν_n suivant

$$\nu_n \left(\mathbb{P}_{\alpha^*, \sigma^*, f^*}^{(n)} \right) = (\theta_2^* \dots \theta_J^*, a_2^* \dots a_J^*, \nu_1^* \dots \nu_J^*).$$

Ce paramètre est différentiable. En effet, pour tout $h = ((\theta, a, v), \sigma, f) \in \mathcal{H}$, on a :

$$\lim_{n \rightarrow \infty} n^{d/2} \left(\nu_n \left(\mathbb{P}_{\alpha_n(h), \sigma_n(h), f_n(h)}^{(n)} \right) - \nu_n \left(\mathbb{P}_{\alpha^*, \sigma^*, f^*}^{(n)} \right) \right) = (\theta, a, v).$$

Par conséquent, il existe une application linéaire continue $\dot{\nu}$ de \mathcal{H} dans $\mathbb{R}^{d(J-1)+2J-1}$. Comme \mathcal{H} est un sous espace d'un espace de Hilbert $\text{cl}(\mathcal{H})$, le théorème de Riesz implique l'existence de $d(J-1) + 2J - 1$ vecteurs $(\dot{\nu}_{j,q}^\theta)_{2 \leq j \leq J, 1 \leq q \leq d}$, $(\dot{\nu}_j^a)_{2 \leq j \leq J}$ et $(\dot{\nu}_j^v)_{1 \leq j \leq J}$ de $\text{cl}(\mathcal{H})$ tel que pour tout $h = ((\theta, a, v), \sigma, f) \in \mathcal{H}$,

$$\begin{aligned} \forall j = 2, \dots, J, \quad \forall q = 1, \dots, d, \quad \langle \dot{\nu}_{j,q}^\theta, h \rangle &= \theta_{j,q}, \quad \langle \dot{\nu}_j^a, h \rangle = a_j, \\ \forall j = 1, \dots, J, \quad \langle \dot{\nu}_j^v, h \rangle &= v_j, \end{aligned}$$

où $\theta = (\theta_{j,q}) \in \mathbb{R}^{d(J-1)}$, $a = (a_j) \in \mathbb{R}^{J-1}$ et $v = (v_j) \in \mathbb{R}^J$. Les calculs explicites des représentants de la différentielle de ν_n sont effectués au lemme 2.10.

Puisque le processus $\Delta_n(h)$ est linéaire en h , on peut lier la notion de linéarité asymptotique et d'efficacité. La proposition suivante est une conséquence de la proposition 5.3 de McNeney et Wellner (2000) (voir le chapitre 1).

Proposition 2.2 (Linéarité asymptotique et efficacité).

Sous les hypothèses de la proposition 2.1, un estimateur asymptotiquement linéaire T_n de ν_n de processus centré,

$$(\Delta_n(\tilde{h}_{21}^\theta) \dots \Delta_n(\tilde{h}_{2q}^\theta), \Delta_n(\tilde{h}_{31}^\theta) \dots \Delta_n(\tilde{h}_{Jq}^\theta), \Delta_n(\tilde{h}_2^a) \dots \Delta_n(\tilde{h}_J^a), \dots \Delta_n(\tilde{h}_J^v)),$$

est régulier efficace si, et seulement si,

$$\begin{aligned} \forall j = 2, \dots, J, \quad \forall q = 1, \dots, d, \quad \tilde{h}_{jq}^\theta &= \dot{\nu}_{jq}^\theta, \quad \tilde{h}_j^a = \dot{\nu}_j^a, \\ \forall j = 1, \dots, J, \quad \tilde{h}_j^v &= \dot{\nu}_j^v. \end{aligned}$$

Plus explicitement, si le modèle est LAN, un estimateur $T_n = (\hat{\theta}_n, \hat{a}_n, \hat{v}_n)$ de α^* est asymptotiquement efficace si :

$$\begin{aligned} n^{d/2} (\hat{\theta}_n - \theta^*) &= \frac{\sigma^*}{n^{d/2}} \sum_{i \in \mathcal{I}_n} \left[\text{Gramm}(\nabla f^*)^{-1} \otimes \left(\frac{1}{a_1^*} \mathbb{I}_{J-1} \mid -D_a^{-1} \right) \right] \nabla F^*(t_i) \epsilon_{i,\cdot} + o_{\mathbb{P}}(1) \\ n^{d/2} (\hat{a}_n - a^*) &= \frac{\sigma^*}{\|f^*\|_{\mathbb{L}^2}} \sum_{i \in \mathcal{I}_n} \left\{ \left[-\frac{a_1^*}{J} V_a \mid I_{J-1} - \frac{1}{J} a a^T \right] F^*(t_i) \right\} \epsilon_{i,\cdot} + o_{\mathbb{P}}(1) \\ n^{d/2} (\hat{v}_n - v^*) &= \sigma^* \sum_{i \in \mathcal{I}_n} \epsilon_{i,\cdot} + o_{\mathbb{P}}(1), \end{aligned}$$

où $\epsilon_i^T = (\epsilon_{i1}, \dots, \epsilon_{iJ})^T$ pour tout $i \in \mathcal{I}_n$, F^* et ∇F^* sont des matrices de dimension respectives $J \times J$ et $(qJ) \times J$, telles que pour tout $t \in \mathbb{R}^d$,

$$F^*(t) = \text{diag}(f^*(t), f^*(t - \theta_2^*) \dots f^*(t - \theta_J^*)),$$

$$\nabla F^*(t) = \begin{pmatrix} \text{diag}(\partial_1 f^*(t), \partial_1 f^*(t - \theta_2^*) \dots \partial_1 f^*(t - \theta_J^*)) \\ \vdots \\ \text{diag}(\partial_d f^*(t), \partial_d f^*(t - \theta_2^*) \dots \partial_d f^*(t - \theta_J^*)) \end{pmatrix},$$

avec $\partial_q f^*$ ($q=1, \dots, d$) la dérivée partielle de f^* suivant la $q^{\text{ème}}$ coordonnée. Les autres notations sont rappelées au théorème 2.4.

On considère maintenant l'estimateur de α^* défini comme le minimum de processus de contraste (2.26). On a vu au théorème 2.3 que cet estimateur est consistant. En fait, il est asymptotiquement efficace.

Théorème 2.4 (Efficacité). *Supposons que les paramètres d'amplitude $(a_j^*)_{j=1 \dots J}$ soient non nuls. On suppose également que la fonction f^* vérifie l'une des deux hypothèses (F1) ou (F2), et qu'elle est différentiable telle que :*

$$\sum_{(l_1, \dots, l_q) \in \mathbb{Z}^d} (|l_1| + \dots + |l_q|) |c_{(l_1, \dots, l_q)}(f^*)| < \infty, \text{ et } \sum_{|l|_\infty > m} |c_l(f^*)| = o\left(\frac{1}{m^{d/2}}\right) \quad (2.32)$$

De plus, m_n vérifie (N1) et (N2) de telle sorte que

$$m_n^{d+1}/n^{d/2} = o(1). \quad (2.33)$$

Alors $(\hat{\theta}_n, \hat{a}_n, \hat{v}_n)$ est asymptotiquement efficace. Les fluctuations $\sqrt{n}(\hat{\theta}_n - \theta^*, \hat{a}_n - a^*, \hat{v}_n - v^*)$ convergent en loi vers le vecteur Gaussien $\mathcal{N}_{d(J-1)+2J-1}(0, \sigma^{*2} H^{-1})$, où H est la matrice inversible définie par

$$H = \begin{pmatrix} \text{Gramm}(\nabla f^*) \otimes (D_a^2 - \frac{1}{J} V_a V_a^T) & 0 & 0 \\ 0 & \|f^*\|_{\mathbb{L}^2}^2 \left(I + \frac{1}{a_1^{*2}} a a^T\right) & 0 \\ 0 & 0 & I_J \end{pmatrix}.$$

et la matrice inverse H^{-1} étant égale à

$$H^{-1} = \begin{pmatrix} (\text{Gramm}(\nabla f^*))^{-1} \otimes \left(D_a^{-2} + \frac{1}{a_1^{*2}} \mathbb{I}_{J-1} \mathbb{I}_{J-1}^T\right) & 0 & 0 \\ 0 & \frac{1}{\|f^*\|_{\mathbb{L}^2}^2} \left(I_{J-1} - \frac{1}{J} a a^T\right) & 0 \\ 0 & 0 & I_J \end{pmatrix}.$$

$\text{Gramm}(\nabla f^*)$ est la matrice de Gramm (annexe A) des dérivées partielles de f^* , $I_{J-1} = \text{diag}(1, \dots, 1)$ et $D_a = \text{diag}(a_2^*, \dots, a_J^*)$ sont des matrices diagonales de taille $J-1$. $\mathbb{I}_{J-1} = (1, \dots, 1)^T$, $a^T = (a_2^*, \dots, a_J^*)^T$ et $V_a = (a_2^{*2}, \dots, a_J^{*2})^T$ sont des vecteurs de \mathbb{R}^{J-1} .

Avant de donner les grandes lignes de la preuve, notons d'abord que si la fonction f^* vérifie **(F1)** ou **(F2)**, alors ses dérivées partielles sont linéairement indépendantes dans l'espace vectoriel $\mathbb{L}^2(\mathcal{D})$. La matrice de Gramm, $\text{Gramm}(\nabla f^*)$, est alors bien inversible (voir le rappel en annexe A). Supposer **(F1)** ou **(F2)** est une hypothèse plus forte d'ailleurs que la simple indépendance linéaire de la famille des dérivées partielles de f^* , puisque en plus elle implique que la fonction est réellement périodique sur le réseau $2\pi\mathbb{Z}^d$.

L'hypothèse $((\sum(|l_1| + \dots + |l_q|)|c_{(l_1, \dots, l_q)}(f^*)| < \infty))$ signifie que les dérivées partielles de f^* sont continues sur \mathbb{R}^d . Elle implique la seconde hypothèse sur la vitesse de convergence de la série de Fourier pour $d = 1$ et $d = 2$. Comme pour le théorème 2.3, l'hypothèse de régularité qu'impose la seconde hypothèse pourrait être contournée en restreignant le nombre de coefficients de Fourier considérés dans le processus de contraste.

D'autre part, on remarque que la matrice des covariances est diagonale par bloc. Les estimateurs des translations $\hat{\theta}_n$, des amplitudes \hat{a}_n , et des niveaux \hat{v}_n sont donc asymptotiquement indépendants. De plus les estimateurs des d vecteurs de \mathbb{R}^J , $(\theta_{1,q}^*, \dots, \theta_{J,q}^*)$ où $q = 1, \dots, d$ sont asymptotiquement indépendants si la matrice de Gramm, $\text{Gramm}(\nabla f^*)$, est diagonale : c'est-à-dire que les dérivées partielles de f^* sont orthogonales dans l'espace $\mathbb{L}^2(\mathcal{D})$. Par exemple pour $d = 2$, la matrice H serait égale à :

$$H = \begin{pmatrix} \|\partial_1 f^*\|_{\mathbb{L}^2}^2 (D_a^2 - \frac{1}{J} V_a V_a^T) & 0 & 0 & 0 \\ 0 & \|\partial_2 f^*\|_{\mathbb{L}^2}^2 (D_a^2 - \frac{1}{J} V_a V_a^T) & 0 & 0 \\ 0 & 0 & \|f^*\|_{\mathbb{L}^2}^2 \left(I + \frac{1}{a_1^{*2}} a a^T\right) & 0 \\ 0 & 0 & 0 & I_J \end{pmatrix}.$$

Ceci montre l'importance du paramétrage de départ des signaux f_j^* étudiés.

Preuve du théorème 2.4. Par définition, le processus de contraste L_n est deux fois différentiable sur \mathcal{A}_3 . Puisque $\hat{\alpha}_n$ est défini comme le minimum de ce processus, on en déduit que la différentielle de L_n s'annule en $\hat{\alpha}_n$. Par un développement de Taylor du gradient de L_n en α^* , il existe $\bar{\alpha}_n \in \mathcal{A}_3$ dans le voisinage de α^* tel que

$$\nabla^2 L_n(\bar{\alpha}_n) n^{d/2}(\hat{\alpha}_n - \alpha^*) = -n^{d/2} \nabla L_n(\alpha^*),$$

où ∇^2 est l'opérateur Hessian. Ainsi, des propositions 2.5 et 2.6, on déduit que :

$$\begin{aligned} n^{d/2}(\hat{\theta}_n - \theta^*) &= \sigma^*(\text{Gramm}(\nabla f^*))^{-1} \otimes \left(D_a^{-2} + \frac{1}{a_1^{*2}} \mathbb{I}_{J-1}\right) \mathbb{G}_n^\theta + o_{\mathbb{P}}(1) \\ n^{d/2}(\hat{a}_n - a^*) &= \frac{\sigma^*}{\|f^*\|_{\mathbb{L}^2}^2} \left(I_{J-1} - \frac{1}{J} a a^T\right) \mathbb{G}_n^a + o_{\mathbb{P}}(1) \\ n^{d/2}(\hat{v}_n - v^*) &= \sigma^* \mathbb{G}_n^v + o_{\mathbb{P}}(1), \end{aligned}$$

où \mathbb{G}_n est défini dans la proposition 2.5. \square

Le vecteur gaussien \mathbb{G}_n , défini dans la preuve du théorème précédent, converge en loi vers $\mathcal{N}_{d(J-1)+2J-1}(0, H)$ de telle sorte que :

$$n^{d/2}(\hat{\alpha}_n - \alpha^*) = (H/\sigma^{*2})^{-1} \sigma^* \mathbb{G}_n + o_{\mathbb{P}}(1).$$

En se référant aux modèles semi-paramétriques indépendants et identiquement distribués (voir chapitre 1), la matrice H/σ^{*2} est la matrice d'information efficace du modèle, et $\sigma^* \mathbb{G}_n$ est le score efficace.

Ce résultat peut être mis en relation avec les résultats de [Murphy et Van der Vaart \(2000\)](#) (voir chapitre 1). Dans cet article, les auteurs y étudient la vraisemblance profilée pour des modèles semi-paramétriques indépendants identiquement distribués. Ils démontrent que si le modèle le moins favorable existe et que le modèle est assez régulier, la vraisemblance profilée se comporte pareillement à la vraisemblance classique, et elle sélectionne correctement la direction la moins favorable. Ceci est vérifié quand la log-vraisemblance profilée admet un développement asymptotique rappelé dans la section 1.2 du chapitre 1. Dans le modèle étudié ici, la log-vraisemblance admet un développement asymptotique similaire. En effet, par la formule de Taylor, il existe $(\tilde{\alpha}_n, \tilde{\sigma}_n^2)$ tel que :

$$\begin{aligned} pl_n(\hat{\alpha}_n, \hat{\sigma}_n) - pl_n(\alpha^*, \sigma^*) &= -\frac{n^d J}{2\sigma^{*2}} \nabla L_n(\alpha^*)(\hat{\alpha}_n - \alpha^*) - \frac{n^d J}{4\sigma^{*2}} (\hat{\alpha}_n - \alpha^*)^T \nabla^2 L_n(\tilde{\alpha}_n)(\hat{\alpha}_n - \alpha^*) + o_{\mathbb{P}}(1). \\ &= n^{d/2} \mathbb{G}_n(\hat{\alpha}_n - \alpha^*) - \frac{n^d}{2} (\hat{\alpha}_n - \alpha^*)^T \frac{H}{\sigma^{*2}} (\hat{\alpha}_n - \alpha^*) + o_{\mathbb{P}}(n^{d/2} \|\hat{\alpha}_n - \alpha^*\| + 1)^2. \end{aligned}$$

Remarque 2.6. Une étude similaire peut être effectuée pour le modèle associé au jeu n^2 de contraintes d'identifiabilité (voir la sous-section 2.1.2). Par exemple, pour $d = 1$, le paramètre à estimer est le vecteur de \mathbb{R}^{3J-3} suivant $(\theta_2^*, \dots, \theta_J^*, a_2^*, \dots, a_J^*, v_2^*, \dots, v_J^*)$. On prouve que l'estimateur défini comme le minimum des moindres carrés profilés est asymptotiquement efficace, mais la matrice de covariance est :

$$\sigma^{*2} \begin{pmatrix} \frac{1}{\|\partial f^*\|_{\mathbb{L}^2}^2} \left(D_a^{-2} + \frac{1}{a_1^{*2}} \mathbb{I}_{J-1} \mathbb{I}_{J-1} \right) & 0 & 0 \\ 0 & \frac{1}{\|f^*\|_{\mathbb{L}^2}^2 - c_0(f^*)^2} B & \frac{-c_0(f^*)}{\|f^*\|_{\mathbb{L}^2}^2 - c_0(f^*)^2} I_{J-1} \\ 0 & \frac{-c_0(f^*)}{\|f^*\|_{\mathbb{L}^2}^2 - c_0(f^*)^2} I_{J-1} & \frac{\|f^*\|_{\mathbb{L}^2}^2}{\|f^*\|_{\mathbb{L}^2}^2 - c_0(f^*)^2} B^{-1} \end{pmatrix}$$

où $B = I_{J-1} - \frac{1}{J} a a^T$ et $B^{-1} = I_{J-1} + \frac{1}{a_1^{*2}} a a^T$. La matrice de covariance pour l'estimateur des translations est identique à celle du théorème 2.4. Par contre, si on se restreint à l'estimation des paramètres d'amplitude et de niveau, la matrice de covariance n'est plus diagonale par blocs : les estimateurs des amplitudes et des niveaux ne

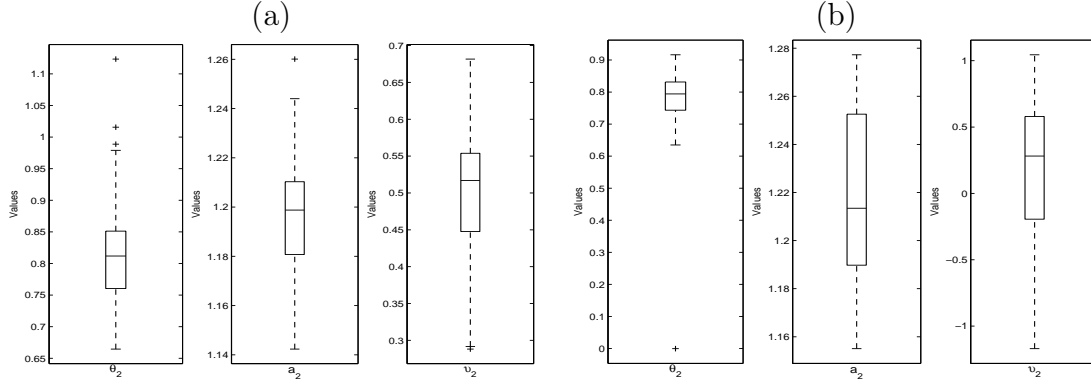


FIG. 2.2: Diagrammes de Tukey (a) pour le jeu n°3 de contraintes d'identifiabilité ($c_0(f^*)=0$), (b) pour le jeu n°2 de contraintes d'identifiabilité ($v_1^*=0$)

sont plus asymptotiquement indépendants. Le coefficient de dépendance entre les deux estimateurs dépend justement du coefficient de Fourier $c_0(f^*)$ qui est fixé à zéro pour le jeu n°3 de contraintes.

L'incidence sur la qualité des estimations est telle que la procédure bootstrap du test d'adéquation développée dans le chapitre 3 (section 3.2) échoue si l'on considère le modèle associé au jeu n°2 de contraintes d'identifiabilité.

Pour illustrer cette conséquence, la figure 2.2 compare les diagrammes en boîte de Tukey des estimateurs associés aux modèles munis des contraintes d'identifiabilité du jeu n°2 et n°3. Les données sont générées à partir des fonctions $g^*(t) = 20(1 - x/(2\pi))x/(2\pi)$ et $f^*(t) = g^*(t) - c_0(g^*)$ ($c_0(g^*) = 10/3$). On estime le paramètre α^* dans chacun des modèles associés au jeu n°2 et n°3. Dans le modèle associé au jeu n°3,

$$\begin{cases} Y_{i1} = a_1^* f^*(t_i) + v_1^* + \epsilon_{i1} & i = 1 \dots n; \\ Y_{i2} = a_2^* f^*(t_i - \theta_2^*) + v_2^* + \epsilon_{i2} & i = 1 \dots n \end{cases};$$

les paramètres à estimer sont les suivants $\theta_2^* = 0.8$, $a_2^* = 1.1990$ (et donc $a_1^* = 0.75$) et $v^* = (a_1^* 10/3 \quad 0.5)$. Dans le modèle équivalent associé au jeu n°2,

$$\begin{cases} Y_{i1} = a_1^* g^*(t_i) + \epsilon_{i1} & i = 1 \dots n \\ Y_{i2} = a_2^* g^*(t_i - \theta_2^*) + (v_2^* - a_2^* 10/3) + \epsilon_{i2} & i = 1 \dots n \end{cases}$$

les paramètres à estimer sont les suivants $\theta_2^* = 0.8$, $a_2^* = 1.1990$ et $v_2^* = 0.5$. Les figures 2.2(a) et 2.2(b) représentent les diagrammes de Tukey pour les estimateurs de θ_2^* , a_2^* et v_2^* pour les modèles associés respectivement au jeu n°3 et n°2.

2.3.3 Estimation de la forme commune f^*

La méthode SEMOR de [Lawton et al. \(1972\)](#) estime simultanément le paramètre α^* et la forme commune f^* . Dans cette sous-section, on étudie la convergence de l'estimateur du maximum de vraisemblance de f^* . Par définition, cet estimateur est la seconde composante de la paire $((\hat{\alpha}_n, \hat{\sigma}_n^2), \hat{f}_n)$ qui maximise $l_n(\alpha, \sigma^2, f)$ sur l'espace des polynômes trigonométriques de degré m_n : c'est donc le polynôme $\hat{f}_n = f_{\hat{\alpha}_n}(\cdot)$, où f_α est le polynôme trigonométrique (2.24) maximisant la vraisemblance pour un $\alpha \in \mathcal{A}_3$ donné.

Avant de discuter de l'éventuelle efficacité de cet estimateur, on étudie d'abord sa consistance pour la norme uniforme des fonctions périodiques et ainsi que pour la norme $\mathbb{L}^2(\mathcal{D})$.

Corollaire 2.1. *Supposons que f^* soit le produit de d fonctions f_1, \dots, f_d définies sur \mathbb{R} k fois continûment différentiables telles que*

$$\begin{aligned} \forall t = (t_1, \dots, t_d) \in \mathbb{R}^d, \quad f(t) &= f(t_1) \dots f_d(t_d) \\ \sum_{l \in \mathbb{Z}^d} (|l_1|^{2k} + \dots + |l_d|^{2k}) |c_l(f^*)|^2 &< \infty \end{aligned} \quad (2.34)$$

et $k \geq (d+1)/2$. De plus, supposons que les hypothèses du théorème 2.4 soient vérifiées, alors il existe une constante C telle que pour n assez grand

$$\begin{aligned} \sup_{t \in \mathbb{R}^d} |\hat{f}_n(t) - f^*(t)| &= \mathcal{O}_{\mathbb{P}} \left(\frac{1}{m_n^{k-1/2}} + \frac{m_n^d}{n^{d/2}} \right), \\ MISE_{f^*}(\hat{f}_n) &= \int_{\mathcal{D}} \mathbb{E} \left(\hat{f}_n(t) - f^*(t) \right)^2 dt \leq C \left(\frac{1}{m_n^{2k}} + \frac{m_n^d}{n^d} \right). \end{aligned}$$

De plus, pour $m_n \sim n^{d/(2k+d)}$, on a $MISE_{f^}(\hat{f}_n) = O(n^{-2dk/(2k+d)})$.*

Pour $d = 1$ on retrouve les vitesses de convergence optimales pour l'estimation non paramétrique ([Tsybakov \(2004\)](#)). Pour $d > 2$, le choix de l'ensemble des coefficients de Fourier $\{l \in \mathbb{Z}^d, |l|_\infty \leq m_n\}$ considéré ici pour estimer la fonction f^* oblige à se restreindre aux fonctions f^* de régularité supérieure à $(d+1)/2$ (afin de satisfaire l'hypothèse 2.28 du théorème 2.4). Si la fonction f^* n'est pas le produit de d fonctions, sous les hypothèses du corollaire (2.1), les mêmes vitesses de convergence sont obtenues pour le risque quadratique intégré. Par contre, la vitesse de convergence pour la norme uniforme est plus faible, car sous l'hypothèse que les dérivées partielles d'ordre k de f^* soient de carré intégrables (2.34) la vitesse de convergence de la série est,

$$\sum_{|l|_\infty > m} |c_l(f^*)| = o(n^{-\tau}), \quad \forall 0 < \tau < k - d/2.$$

On étudie maintenant l'efficacité de l'estimateur \hat{f}_n . Soit $\text{cl}(\mathcal{F}_0)$ la fermeture de \mathcal{F}_0 pour la norme $\mathbb{L}^2(\mathcal{D})$,

$$\text{cl}(\mathcal{F}_0) = \{f \in \mathbb{L}^2(\mathcal{D}) \text{ telle que } c_0(f) = 0\}.$$

La suite de paramètres ν_n étudié est la série de Fourier tronquée de f^* ,

$$\nu_n(\mathbb{P}_{\alpha^*, f^*}^n) = \sum_{|l|_\infty \leq m_n} c_l(f^*) e^{il(\cdot)}.$$

Si f^* est une fonction de carré intégrable, cette suite de paramètres est différentiable,

$$\lim_{n \rightarrow \infty} \sqrt{n} \left(\nu_n \left(\mathbb{P}_{\alpha_n(h), \sigma_n(h)^2, f_n(h)}^{(n)} \right) - \nu_n \left(\mathbb{P}_{\alpha^*, \sigma^{*2}, f^*}^{(n)} \right) \right) = \dot{\nu}(h) = f.$$

La suite de paramètres est à valeurs dans l'espace de Hilbert $\text{cl}(\mathcal{F}_0)$. L'espace dual $\text{cl}(\mathcal{F}_0)^*$ de $\text{cl}(\mathcal{F}_0)$ permet alors d'obtenir une représentation de $\dot{\nu}$ dans $\text{cl}(\mathcal{H})$, (comme dans la sous-section 2.3.2 où la suite des paramètres était à valeurs dans un espace de dimension finie). En effet par le théorème de Riesz, il existe pour tout $b^* \in \mathbb{B}^*$ un vecteur $\dot{\nu}_{b^*} \in \text{cl}(\mathcal{H})$ tel que

$$\forall h = (\theta, a, \sigma^2, f) \in \mathcal{H} \quad b^* \dot{\nu}(h) = \langle \dot{\nu}_{b^*}, h \rangle = b^* f.$$

L'espace $\text{cl}(\mathcal{F}_0)$ étant un espace de Hilbert, l'espace dual $\text{cl}(\mathcal{F}_0)^*$ est engendré par la famille suivante de formes linéaires,

$$b_{1l}^* : f \in \mathcal{F}_0 \rightarrow \int_{\mathcal{D}} f(t) \cos(l \cdot t) \frac{dt}{2\pi^d} \quad \text{et} \quad b_{2l}^* : f \in \mathcal{F}_0 \rightarrow \int_{\mathcal{D}} f(t) \sin(l \cdot t) \frac{dt}{2\pi^d}, \quad l \in \mathbb{Z}^d \setminus \{0\}.$$

La connaissance des vecteurs $\dot{\nu}_{b_{1l}^*}$ et $\dot{\nu}_{b_{2l}^*}$ pour tout $l \in \mathbb{Z}^d \setminus \{0_d\}$ suffit donc à déterminer tous les autres vecteurs $\{\dot{\nu}_{b^*}, b^* \in \text{cl}(\mathcal{F}_0)^*\}$. Après quelques calculs, on a pour $t \in \mathbb{R}^d$,

$$\dot{\nu}_{b_{1l}^*}(t) = \left((0, 0, 0), 0, \sigma^{*2} \frac{\cos(l \cdot t)}{J} \right) \quad \text{et} \quad \dot{\nu}_{b_{2l}^*}(t) = \left((0, 0, 0), 0, \sigma^{*2} \frac{\sin(l \cdot t)}{J} \right).$$

Le lemme 2.12 établit que l'estimateur \hat{f}_n est asymptotiquement linéaire, au sens de la définition 1.5.

Or d'après la proposition 1.2 de McNeney et Wellner (2000), un pré-requis pour que l'estimateur \hat{f}_n soit éventuellement efficace est que pour toute forme linéaire b^* de $\text{cl}(\mathcal{F}_0)$, $b^*(\hat{f}_n)$ soit un estimateur efficace de $b^*(f^*)$.

Corollaire 2.2. *$b^*(\hat{f}_n)$ est un estimateur asymptotiquement régulier efficace de $b^*(f^*)$ pour tout $b^* \in \text{cl}(\mathcal{F}_0)^*$ si, et seulement si f^* est une fonction paire ou impaire. En particulier, les estimateurs des coefficients de Fourier de f^* sont asymptotiquement efficaces.*

L'estimateur de f^* peut alors être efficace, si la fonction est paire ou impaire. Admettons que cette dernière hypothèse soit effectivement vérifiée par la forme commune f^* du modèle. L'estimateur \hat{f}_n serait asymptotiquement efficace si, et seulement si pour toute direction $h \in \mathcal{H}$ les fluctuations $n^{d/2} \left(\hat{f}_n - \nu_n \left(\mathbb{P}_{\alpha_n(h), \sigma_n(h)^2, f_n(h)}^{(n)} \right) \right)$ convergèrent faiblement sous $\mathbb{P}_{\alpha_n(h), \sigma_n(h)^2, f_n(h)}^{(n)}$ dans $\text{cl}(\mathcal{F}_0)$. Or pour la direction $h = ((0, 0, 0), 0, 0)$, $\sqrt{n} \left(T_n - \nu_n \left(\mathbb{P}_{\alpha^*, f^*}^{(n)} \right) \right)$ diverge sous $\mathbb{P}_{\alpha^*, f^*}^{(n)}$ (d'après le corollaire 2.1). L'estimateur \hat{f}_n n'est donc pas asymptotiquement efficace.

Remarque 2.7. *Le modèle où la fonction f^* est supposée symétrique (ou antisymétrique) est étudié par Dalalyan (2005). Dans un tel modèle, la condition d'identifiabilité sur le paramètre de translation n'a aucune raison d'être, puisque par définition l'origine est un repère pour l'estimer. Dans ce modèle, les estimateurs de $\theta_1^*, \dots, \theta_j^*$ seraient, par conséquent, asymptotiquement indépendants. De plus cette estimation serait adaptative pour les paramètres de translation.*

2.4 Application des méthodes d'estimation

Dans les section précédentes, nous avons proposé deux critères d'estimation pour le paramètre α^* . Le critère W_n (2.16) est défini pour le modèle associé au jeu n°1 de contraintes d'identifiabilité et utilise une suite de poids. Le critère L_n (2.26) est associé au jeu n°3 de contraintes d'identifiabilité, il contrôle le bruit en limitant le nombre m_n de coefficients de Fourier. Dans cette section, nous discutons du choix des contraintes d'identifiabilité, et du choix de la suite de poids ou du nombre de coefficients de Fourier.

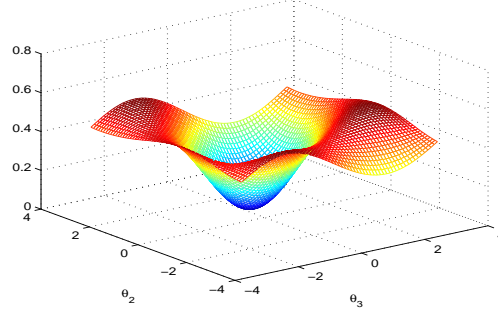
2.4.1 Les contraintes d'identifiabilité

On propose d'étudier l'effet des contraintes d'identifiabilité sur l'estimation. Les simulations présentées sont générées par MATLAB. Notre but est de comparer les fonctions de contrastes associées aux processus W_n et L_n à partir de deux sous-modèles. Dans le premier, les fonctions de régression f_j^* ne diffèrent que d'un paramètre de translation θ_j^* , alors que pour le second sous-modèle c'est seulement en amplitude que les fonctions f_j^* diffèrent.

Les données sont générées à partir de la fonction 2π -périodique g^* suivante :

$$g^*(x) = 20 * \{1 - x/(2\pi)\} x/(2\pi), \quad x \in [0, 2\pi],$$

(ici $c_l(g^*) = 8\pi^2/l^2$ pour $l \neq 0$). L'écart type étant pris égal à 1, le rapport signal sur bruit est alors de 3. Afin de comparer les critères W_n et L_n sur les contraintes

FIG. 2.3: Modèle de translation, $J = 3$. Graphique de W_n .

d'identifiabilité, on choisit les paramètres suivants :

$$n = 125, \quad m_n = 5 \quad \text{et} \quad \delta_l = \begin{cases} 1 & |l| \leq 5, \\ 0 & |l| > 5. \end{cases}$$

Les estimateurs sont calculés à partir de la procédure d'optimisation `fminsearch` de MATLAB. Cet algorithme utilise la méthode simplexe de [Lagarias et al. \(1999\)](#). Les programmes sont assez simple, et sont donnés en annexe.

Modèle de translations pour $J = 3$ sujets. Le modèle considéré est le suivant,

$$\begin{cases} Y_{i1} = g(t_i) + \epsilon_{i1} & i = 1 \dots n \\ Y_{i2} = g(t_i - \theta_2^*) + \epsilon_{i2} & i = 1 \dots n \\ Y_{i3} = g(t_i - \theta_3^*) + \epsilon_{i3} & i = 1 \dots n \end{cases}$$

avec $\theta^* = (0 \ 0.95 \ 0.25)$. On souhaite estimer les paramètres de translation θ_2^* et θ_3^* à partir des données (Y_{ij}) seules. Le processus de contraste utilisé est identique à (2.16), sauf que les coefficients de Fourier (2.14) sont uniquement rephasés :

$$\tilde{c}_{jl}(\theta) = d_{jl} e^{il\theta_j}, \quad j = 1 \dots J, \ |l| < n/2.$$

Pour $n = 125$, le minimum du contraste est atteint en $\hat{\theta}_n = (0 \ 0.9247 \ 0.2847)$. Une visualisation de ce processus de contraste est donnée par la figure 2.3. Ce qui est remarquable, c'est que le minimum de contraste est atteint dans une sorte de "puits". (Puisque le processus dérivée de L_n pour ce sous-modèle est identique à une constante additive près, il n'est pas considéré ici.)

Modèle d'amplitude pour $J = 3$ sujets. Le modèle considéré est le suivant,

$$\begin{cases} Y_{i1} = a_1^* g(t_i) + \epsilon_{i1} & i = 1 \dots n \\ Y_{i2} = a_2^* g(t_i) + \epsilon_{i2} & i = 1 \dots n \\ Y_{i3} = a_3^* g(t_i) + \epsilon_{i3} & i = 1 \dots n \end{cases}$$

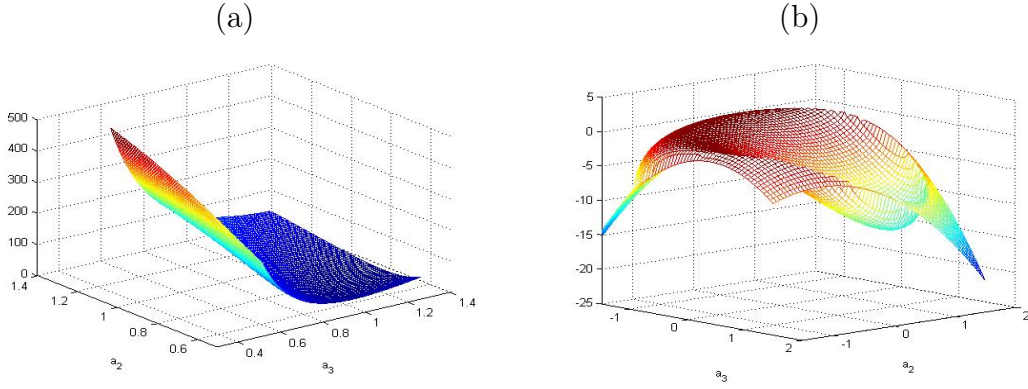


FIG. 2.4: Modèle d'amplitude, $J = 3$: (a) graphique de W_n , (b) graphique de L_n

avec $a^* = (1 \ 0.75 \ 1.199)$. On souhaite estimer les paramètres d'amplitude a_2^* et a_3^* à partir des données (Y_{ij}) seules. Considérons d'abord le processus dérivé de W_n pour ce sous-modèle muni du jeu n°1 de contraintes d'identifiabilité. Il est défini de manière identique à (2.16), sauf que les coefficients de Fourier (2.14) sont :

$$\tilde{c}_{jl}(\theta) = d_{jl}/a_j, \quad j = 1 \dots J, \quad |l| < n/2.$$

Pour $n = 125$, le minimum du contraste est atteint en $\hat{a}_n = (1 \ 0.7950 \ 1.2364)$. Une visualisation de ce processus de contraste est donnée par la figure 2.4(a).

Concernant le processus de contraste dérivé de L_n , il est défini de façon suivante,

$$L_n(a) = \frac{1}{nJ} \sum_{j=1}^J \sum_{i=1}^n (Y_{ij})^2 - \sum_{1 \leq |l| \leq m_n} \left| \frac{1}{J} \sum_{j=1}^J a_j d_{j,l} \right|^2.$$

Pour $n = 125$, le minimum du contraste est atteint en $\hat{a}_n = (1.0040 \ 0.7657 \ 1.1856)$. Une visualisation de ce processus de contraste est donnée par la figure 2.4(b).

La fonction associée à la contrainte d'identifiabilité ($a_1 = 1$) est plus aplatie (figure 2.4(a)). Ainsi la contrainte d'identifiabilité ($\sum_{j=1}^J a_j^2 = J$) définit une fonction de contraste plus convexe : son minimum global est plus visible graphiquement.

2.4.2 Choix de la suite de poids.

Nous discutons du choix de la suite de poids à partir du modèle décrit à la sous-section 2.2.2 pour $J = 2$ unités :

$$\begin{cases} Y_{i1} = f^*(t_i) + \epsilon_{i1} & i = 1 \dots n \\ Y_{i2} = a_2^* f^*(t_i - \theta_2^*) + \epsilon_{i2} & i = 1 \dots n \end{cases}$$

où $\theta^* = (0 \ 0.95)$ et $a^* = (1 \ 0.75)$. On souhaite estimer les paramètres de translation θ_2^* et d'amplitude a_2^* à partir des données (Y_{ij}) et des temps d'observation (t_i) . La procédure d'estimation est celle décrite dans la sous-section 2.2.2 pour différentes suites de poids. Le tableau 2.1 présente les moyennes et les écarts-type pour les estimateurs de θ_2^* et a_2^* pour trois suites de poids. En comparant les différentes colonnes du tableau

	Poids n°1 $\delta_l = 1/(l^2 + 1),$ $\forall l \in \mathbb{Z}.$	Poids n°2 $\delta_l = 1, \text{ si } l \leq 5$ $\delta_l = 0 \text{ sinon}$	Poids n°3 $\delta_l = 1, \text{ si } l \leq 15$ $\delta_l = 0 \text{ sinon}$
$n = 31$	$\hat{\theta} = 0.9538 \ (0.2023)$ $\hat{a} = 0.7556 \ (0.0655)$	$\hat{\theta} = 0.9230 \ (0.2472)$ $\hat{a} = 0.7784 \ (0.0664)$	$\hat{\theta} = 0.3470 \ (0.7144)$ $\hat{a} = 0.8690 \ (0.1415)$
$n = 51$	$\hat{\theta} = 0.9545 \ (0.1623)$ $\hat{a} = 0.7528 \ (0.0493)$	$\hat{\theta} = 0.9460 \ (0.1612)$ $\hat{a} = 0.7688 \ (0.0522)$	$\hat{\theta} = 0.5722 \ (0.5519)$ $\hat{a} = 0.8216 \ (0.0925)$
$n = 101$	$\hat{\theta} = 0.9498 \ (0.7509)$ $\hat{a} = 0.7509 \ (0.0364)$	$\hat{\theta} = 0.9454 \ (0.1055)$ $\hat{a} = 0.7587 \ (0.0352)$	$\hat{\theta} = 0.8828 \ (0.2385)$ $\hat{a} = 0.7788 \ (0.0459)$
$n = 201$	$\hat{\theta} = 0.9524 \ (0.0817)$ $\hat{a} = 0.7496 \ (0.0255)$	$\hat{\theta} = 0.9496 \ (0.0706)$ $\hat{a} = 0.7549 \ (0.0237)$	$\hat{\theta} = 0.9499 \ (0.1086)$ $\hat{a} = 0.7633 \ (0.0274)$

TAB. 2.1: Moyennes et écarts type (entre parenthèses) pour l'estimation des paramètres $\theta_2^* = 0.95$ et $a_2^* = 0.75$ calculés à partir de 1000 jeux de données.

2.1, on constate tout d'abord qu'il n'est pas nécessaire de considérer beaucoup de coefficients de Fourier pour avoir une bonne estimation des paramètres. Au contraire quand l'échantillon est petit il est préférable d'en considérer le moins possible (cf. les colonnes associées au poids n°2 et n°3). Remarquons par ailleurs qu'il suffit d'avoir au plus deux coefficients de Fourier non nuls (cf. les hypothèses (W1) ou (W2)) pour que l'estimateur soit consistant (théorème 2.1).

La colonne associée au poids n°1 illustre que le rôle de la suite de poids est bien de contrôler la partie stochastique du bruit. Quelque soit la taille de l'échantillon, la qualité de l'estimation reste à peu près identique à celle où l'on ne considère que cinq coefficients de Fourier.

En conclusion, le choix de la suite de poids est un compromis entre le niveau de bruit et la taille des données. Plus le niveau de bruit est grand, plus il est préférable de prendre une suite de poids régulière ou très concentrée sur un petit nombre de coefficients de Fourier.

2.4.3 Choix du nombre de coefficients de Fourier m_n

Nous venons de voir qu'il n'est pas dans notre intérêt de choisir m_n trop grand dans le processus M_n . En effet les théorèmes 2.3 et 2.4 exigent que m_n soit relativement petit par rapport à la taille n de l'échantillon. Nous proposons de discuter du choix de m_n pour le sous-modèle de translation,

$$Y_{i,j} = f^*(t_i - \theta_j^*) + \sigma^* \epsilon_{i,j} \quad i = 1 \dots n, \quad j = 1 \dots J.$$

On se restreint au cas de deux courbes ($J = 2$). Avec les contraintes d'identifiabilité posées sur la modèle ($\theta_1^* = 0$), le paramètre à estimer est θ_2^* . Dans la suite de l'exposé, on suppose que $m_n = m \geq 1$ est fixé. Nous cherchons à avoir une idée sur comment choisir le nombre m le plus optimal.

L'estimateur des translations $\hat{\theta}_{n,m}$ minimise le processus des moindres carrés profilé :

$$L_{n,m}(\theta) = \frac{1}{nJ} \sum_{j=1}^J \sum_{i=1}^n (Y_{ij})^2 - \sum_{1 \leq |l| \leq m} |\hat{c}_l(\theta)|^2,$$

où $\hat{c}_l(\theta) = \sum_{j=1}^J d_{j,l} e^{il\theta_j} / J$ et $\theta = (0, \theta_2)$. Afin que l'une des deux hypothèses **(F1)** ou **(F2)** soit vérifiée, on suppose que le coefficient de Fourier $c_1(f^*)$ est non nul. Pour simplifier l'étude, on substitue dans le modèle (2.11) les coefficients de Fourier discrets par les coefficients de Fourier des fonctions f_j^* , (cette simplification a été considérée dans la sous-section 2.2.2),

$$d_{j,l} = c_l(f^*) e^{-il\theta_j^*} + \sigma^* w_{j,l}, \quad j = 1, \dots, J, \quad |l| \leq m.$$

Les théorèmes 2.1 et 2.2 permettent d'établir respectivement la consistance et la normalité asymptotique de l'estimateur par minimum de contraste $L_{n,m}$. Le risque quadratique asymptotique est alors égal à :

$$\frac{2\sigma^{*2}}{n \sum_{|l| \leq m} l^2 |c_l(f)|^2}.$$

Ainsi, ce risque quadratique asymptotique de l'estimateur diminue quand m augmente. Or dans le tableau 2.1 et pour des échantillon pas trop grand, on observe le contraire. Il existe donc un terme de biais négligeable si la taille de l'échantillon est grande, mais déterminant pour le choix de m . Nous donnons ici l'heuristique pour calculer l'influence dans le risque quadratique de l'estimateur de ce terme de biais pour n fixé. En reprenant la preuve du théorème 2.4, il existe un point $\tilde{\theta}_{n,m}$ dans un proche voisinage de θ^* tel que :

$$\nabla^2 L_{n,m}(\tilde{\theta}_{n,m})(\hat{\theta}_n - \theta^*) = \nabla L_{n,m}(\theta^*).$$

Comme $\nabla^2 L_{n,m}(\tilde{\theta}_{n,m})$ converge en probabilité vers $\sum_{|l| \leq m} l^2 |c_l(f^*)|^2 / 2$, (voir la proposition 2.6), on identifie ce terme à sa limite. D'autre part, en reprenant les notations de la preuve de la proposition 2.5, $\nabla L_{n,m}(\theta^*)$ est la somme de deux termes decorrélés :

$$\nabla L_{n,m}(\theta^*) = \sigma^* \nabla G_{n,m}(\theta^*) + \sigma^{*2} \nabla Q_{n,m}(\theta^*).$$

Après calculs, la variance de $\nabla L_{n,m}$ est :

$$\frac{\sigma^{*2}}{2n} \left(\sum_{|l| \leq m} l^2 |c_l(f^*)|^2 \right) + \sigma^{*4} \frac{m(m+1)(2m+1)}{12n^2}.$$

Par conséquent, une approximation du risque quadratique de l'estimateur de θ_2^* est :

$$R(m) = \frac{2\sigma^{*2}}{n} \left(\sum_{|l| \leq m} l^2 |c_l(f^*)|^2 \right)^{-1} + \sigma^{*4} \frac{m(m+1)(2m+1)}{3n^2} \left(\sum_{|l| \leq m} l^2 |c_l(f^*)|^2 \right)^{-2}.$$

Le premier terme de la somme est le risque quadratique asymptotique. Le second terme provient du biais pour l'estimation des coefficients de Fourier de $l_{c_l}(f^*)$. Ainsi on s'aperçoit que la choix de m doit prendre en compte la taille de l'échantillon n , la régularité de la fonction f^* et le rapport signal sur bruit.

2.5 Preuves de la Section 2.2

Orthogonalité du bruit complexe (w_{jl}).

Nous proposons tout d'abord une étude sur le comportement asymptotique des bruits complexes, notés $(w_{j,l})$. Ces bruits apparaissent dans la transformée de Fourier discrète des données (2.11). Cette étude est menée dans le cas où les variables $(\epsilon_{i,j})$ sont indépendants identiquement distribuées (elles ne sont pas supposées gaussiennes). Le lecteur peut se reporter à l'annexe A pour un rappel sur les variables gaussiennes complexes. Dans la suite de l'exposé, les variables considérées sont les variables normalisées,

$$\xi_{j,l} = n^{d/2} w_{j,l}.$$

C'est donc le comportement asymptotique de ces dernières variables que nous étudions.

Remarquons d'abord que pour tout $l \in \mathbb{Z}^d$ fixé et pour n assez grand (tel que $|l|_\infty < n/2$), nous avons les relations suivantes :

$$\Re(\xi_{j,l}) = \Re(\xi_{j,-l}) \quad \text{et} \quad \Im(\xi_{j,l}) = -\Im(\xi_{j,-l}). \quad (2.35)$$

On est alors amené à étudier la famille de variables suivante :

$$\mathcal{X}_n = \{\xi_{j,l}, j = 1 \dots J, l = (l_1, \dots, l_d) \in \mathbb{Z}^d \text{ tel que } 2|l|_\infty < n \text{ et } l_1 > 0\}.$$

Lemme 2.1. *La famille \mathcal{X}_n converge en loi vers une famille de variables gaussiennes indépendantes complexes centrées réduites $\mathcal{N}^{\mathbb{C}}(0, 1)$.*

Preuve du lemme 2.1. Soient $j \in \{1 \dots J\}$ et $l = (l_1, \dots, l_d) \in \mathbb{Z}^d$ tel que $l \neq 0_d$ et $l_1 > 0$ fixés. Pour n assez grand, la variable $\xi_{j,l}$ est bien définie et appartient à la famille \mathcal{X}_n . Notons $\xi_{j,l}^x$ la partie réelle de $\xi_{j,l}$, et $\xi_{j,l}^y$ sa partie imaginaire. On montre d'abord que le vecteur $(\xi_{j,l}^x, \xi_{j,l}^y)$ converge en loi vers $\mathcal{N}_2(0, I_2/2)$. En effet, l'orthogonalité de la base de Fourier discrète (2.9) permet de déduire que la matrice de covariance de $(\xi_{j,l}^x, \xi_{j,l}^y)$ est $I_2/2$. Ainsi pour tout $(u_1, u_2) \in \mathbb{R}^2$ et pour tout $\eta > 0$, on a :

$$\begin{aligned} \text{var}(u_1 \xi_{j,l}^x + u_2 \xi_{j,l}^y) &= (u_1^2 + u_2^2)/2 \\ \sum_{i \in \mathcal{I}_n} \frac{(u_1 \cos(l \cdot t_i) + u_2 \sin(l \cdot t_i))^2}{n^d} &\mathbb{P} \left(|(u_1 \cos(l \cdot t_i) + u_2 \sin(l \cdot t_i)) \epsilon_{i,j}| > n^{d/2} \eta \right) \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Le théorème de Lindeberg-Feller (cf. l'annexe A) permet de conclure.

De façon similaire, on montre que la famille de variables \mathcal{X}_n converge en loi vers une famille de variables gaussiennes complexes indépendantes $\mathcal{N}^{\mathbb{C}}(0, 1)$. Soit $c = (c_{j,l})$ une suite ayant un nombre fini de termes non nuls tels que pour tout $l \in \mathbb{Z}^d$ et pour tout $j = 1, \dots, J$ $c_{j,l} = \overline{c_{j,-l}}$. On souhaite montrer que la variable $\sum_{j,l} c_{j,l} \xi_{j,l}$ converge en distribution vers une Gaussienne complexe centrée de variance $\sum_{j,l} |c_{j,l}|^2$. Il suffit de remarquer que pour n assez grand, cette variable peut se récrire sous la forme suivante,

$$\sum_{j,l} c_{j,l} \xi_{j,l} = \frac{1}{n^{d/2}} \sum_{j=1}^J \sum_{r \in \mathcal{I}_n} \left(\sum_{\substack{2|l|_{\infty} \leq n \\ l_1 > 0}} c_{j,l} e^{-il \cdot t_r} \right) \epsilon_{r,j}.$$

En utilisant l'orthogonalité de la base de Fourier discrète et le théorème de Lindeberg-Feller, le lemme est prouvé. \square

Lemmes techniques pour la preuve du théorème 2.1

Les deux lemmes qui suivent sont utilisés dans la preuve du théorème sur la consistance de l'estimateur $(\hat{\theta}_n, \hat{a}_n)$ défini comme le minimum du processus de contraste W_n .

Lemme 2.2. *Sous les mêmes hypothèses que le théorème 2.1, le processus de contraste W_n converge uniformément en probabilité vers la fonction de contraste K*

$$K(\theta, a) = \frac{1}{J} \sum_{j=1}^J |a_j^*/a_j|^2 \int_0^{2\pi} |\psi * f(t)|^2 \frac{dt}{2\pi} - \int_0^{2\pi} \left| \frac{1}{J} \sum_{j=1}^J \frac{a_j^*}{a_j} \psi * f(t + \theta_j - \theta_j^*) \right|^2 \frac{dt}{2\pi}.$$

Démonstration. Remarquons d'abord que :

$$\frac{1}{J} \sum_{j=1}^J \Re(\tilde{c}_{jl}(\alpha) \bar{\tilde{c}}_l(\alpha)) = |\hat{c}_l(\alpha)|^2. \quad (2.36)$$

On peut alors récrire le processus de contraste W_n sous la forme,

$$W_n(\theta, a) = \frac{1}{J} \sum_{j=1}^J \sum_{l=-(n-1)/2}^{(n-1)/2} \delta_l^2 (|\tilde{c}_{jl}(\alpha)|^2 - |\hat{c}_l(\alpha)|^2).$$

Posons $C(a, a^*) = J^{-1} \sum_{j=1}^J |a_j^*/a_j|^2$, et $\gamma_l = \arg(c_l(f))$. En utilisant les notations introduites au lemme 2.1, le processus est la somme des trois termes D_n , L_n/\sqrt{n} et Q_n/n définis ci après :

$$\begin{aligned} D_n(\theta, a) &= C(a, a^*) \sum_{l=-(n-1)/2}^{(n-1)/2} \delta_l^2 |c_l(f)|^2 - \sum_{l=-(n-1)/2}^{(n-1)/2} \left| \frac{1}{J} \sum_{j=1}^J \delta_l c_l(f) \frac{a_j^*}{a_j} e^{il(\theta_j - \theta_j^*)} \right|^2 \\ L_n(\theta, a) &= \frac{2(J-1)}{J^2} \sum_{l=-(n-1)/2}^{(n-1)/2} \delta_l^2 |c_l(f)| \sum_{j=1}^J \frac{a_j^*}{a_j^2} \left(\cos(\gamma_l - l\theta_j^*) \xi_{j,l}^x + \sin(\gamma_l - l\theta_j^*) \xi_{j,l}^y \right) \\ &\quad - \frac{2}{J^2} \sum_{2|l \leq n-1} \sum_{\substack{j=1 \dots J \\ k \neq j}} \frac{a_j^* \delta_l^2 |c_l(f)|}{a_j a_k} \left\{ \cos(l(\theta_j - \theta_j^* - \theta_k) + \gamma_l) \xi_{k,l}^x + \sin(l(\theta_j - \theta_j^* - \theta_k) + \gamma_l) \xi_{k,l}^y \right\}, \\ Q_n(\theta, a) &= \frac{J-1}{J^2} \sum_{l=-(n-1)/2}^{(n-1)/2} \delta_l^2 \sum_{j=1}^J \frac{(\xi_{j,l}^x)^2 + (\xi_{j,l}^y)^2}{|a_j|^2} \\ &\quad - \frac{2}{J^2} \sum_{l=-(n-1)/2}^{(n-1)/2} \delta_l^2 \sum_{j=1}^J \sum_{k>j} \frac{1}{a_j a_k} \left\{ \cos l(\theta_j - \theta_k) (\xi_{j,l}^x \xi_{k,l}^x + \xi_{j,l}^y \xi_{k,l}^y) + \sin l(\theta_j - \theta_k) (\xi_{j,l}^x \xi_{k,l}^y - \xi_{j,l}^y \xi_{k,l}^x) \right\} \end{aligned}$$

On étudie séparément l'uniforme convergence de ces termes.

- Les séries (δ_l) et $(c_l(f^*))$ sont de carré sommable d'après l'hypothèse (2.18). Ainsi, par le théorème de Parseval, $D_n(\theta, a)$ converge uniformément vers,

$$C(a, a^*) \int_0^{2\pi} |\psi * f(t)|^2 \frac{dt}{2\pi} - \int_0^{2\pi} \left| \frac{1}{J} \sum_{j=1}^J \frac{a_j^*}{a_j} \psi * f(t + \theta_j - \theta_j^*) \right|^2 \frac{dt}{2\pi},$$

où ψ est la \mathbb{L}^2 -fonction (2.17) associée à la suite de poids et $*$ est l'opérateur de convolution.

- Le terme Q_n est une forme quadratique des variables complexes $(\xi_{j,l})$. Il est borné à une constante près par la variable positive,

$$Q^B = \sum_{l=-(n-1)/2}^{(n-1)/2} \delta_l^2 \sum_{j=1}^J \{(\xi_{j,l}^x)^2 + (\xi_{j,l}^y)^2\}.$$

Puisque la série (δ_l) est de carré sommable, et par le lemme 2.1, l'espérance mathématique de Q_B est bornée. Ainsi Q_n/n converge uniformément en probabilité vers zéro d'après l'inégalité de Markov.

- D'après les contraintes d'identifiabilité posées sur le modèle, on a $0 < a_{\min} \leq a_j \leq a_{\max}$. Par conséquent, L_n est bornée à une constante près par

$$L^B = \sum_{l=-(n-1)/2}^{(n-1)/2} \delta_l^2 |c_l(f)| \sum_{j=1}^J (|\xi_{j,l}^x| + |\xi_{j,l}^y|).$$

Ainsi, d'après (2.18) et le lemme (2.1), on a :

$$\mathbb{E}(L^B) \leq \sum_{l \in \mathbb{Z}} \delta_l^2 |c_l(f^*)| J \sqrt{2} < \infty.$$

Par conséquent L^B/\sqrt{n} converge en probabilité vers 0. Et donc L_n/\sqrt{n} converge uniformément en probabilité vers 0.

□

Lemme 2.3 (Unicité du minimum). *K admet un unique minimum en (θ^*, a^*) .*

Démonstration. D'après l'inégalité de Cauchy-Schwarz, la fonction K est à valeur positive. Par définition, K s'annule en (θ^*, a^*) . Soit (θ, a) un minimum de K , c'est à dire $K(\theta, a) = 0$. Le point (θ, a) vérifie donc le cas d'égalité dans l'inégalité de Cauchy-Schwarz,

$$C(a, a^*) \sum_{l \in \mathbb{Z}} \delta_l^2 |c_l(f)|^2 = \sum_{l \in \mathbb{Z}} \delta_l^2 |c_l(f)|^2 \left| \frac{1}{J} \sum_{j=1}^J \frac{a_j^*}{a_j} e^{il(\theta_j - \theta_j^*)} \right|^2 \quad (2.37)$$

Or, pour $l \in \mathbb{Z}$ fixé, on a par l'inégalité de Cauchy-Schwarz,

$$\left| \frac{1}{J} \sum_{j=1}^J \frac{a_j^*}{a_j} e^{il(\theta_j - \theta_j^*)} \right|^2 \leq C(a, a^*). \quad (2.38)$$

Comme la suite (δ_l^2) est positive, l'équation (2.37) est vraie si, et seulement si les inégalités (2.38) sont des égalités pour tout $l \in \mathbb{Z}$ tel que $\delta_l^2 |c_l(f)|^2 \neq 0$. Considérons un tel entier relatif l ($\delta_l^2 |c_l(f)|^2 \neq 0$). Il existe donc un nombre complexe $\lambda_l \in \mathbb{C}$ tel que les deux vecteurs de \mathbb{C}^J , $(a_j^*/a_j)_j$ et $(e^{-il(\theta_j - \theta_j^*)})_j$, soient colinéaires :

$$a_j^*/a_j = \lambda_l e^{-il(\theta_j - \theta_j^*)}, \quad \forall j = 1, \dots, J.$$

Par les contraintes d'identifiabilité posées sur le modèle ($a_1 = a_1^* = 1$ et $\theta_1 = \theta_1^* = 0$), λ_l est égal à 1, et donc $a = a^*$ et

$$l(\theta_j - \theta_j^*) \equiv 0(2\pi), \quad \forall j = 1 \dots J.$$

Si **(W1)** est vérifiée, alors pour $l = \pm 1$, on a bien que $\theta = \theta^*$. Si **(W2)** est vérifiée, alors il existe p et l deux entiers premiers entre eux tels que $l(\theta_j - \theta_j^*) \equiv 0(2\pi)$ et $p(\theta_j - \theta_j^*) \equiv 0(2\pi)$. Or, par définition p et l sont premiers entre eux, il existe donc deux entiers u et v tel que $ul + vp = 1$. On déduit que $(\theta_j - \theta_j^*) \equiv 0(2\pi)$, et donc $\theta = \theta^*$. \square

Lemme et propositions pour la preuve du théorème 2.2

Les propositions 2.3 et 2.4, relatives à l'étude de la convergence du gradient et de la matrice hessienne de W_n , nécessitent le calculs des dérivées partielles. Ces calculs sont basés sur le lemme 2.4.

Lemme 2.4. *Soient k et m deux entiers tel que $2 \leq k, m \leq J$, $k \neq m$. Alors les dérivées premières et secondes de la fonction W_n (2.16) en $(\theta, a) \in \mathcal{A}_1$ sont*

$$\begin{aligned} \frac{\partial W_n}{\partial \theta_k}(\alpha) &= \frac{2}{J} \sum_{l=-(n-1)/2}^{(n-1)/2} \delta_l^2 l \Im(\tilde{c}_{k,l} \bar{\tilde{c}}_l), & \frac{\partial W_n}{\partial a_k}(\alpha) &= \frac{2}{J a_k} \sum_{l=-(n-1)/2}^{(n-1)/2} \delta_l^2 \Re(\tilde{c}_{k,l} \bar{\tilde{c}}_l - |\tilde{c}_{k,l}|^2). \\ \frac{\partial^2 M_n}{\partial \theta_k \partial \theta_m}(\alpha) &= \frac{-2}{J^2} \sum_{|2l| \leq n-1} \delta_l^2 l^2 \Re(\tilde{c}_{k,l} \bar{\tilde{c}}_l), & \frac{\partial^2 W_n}{\partial \theta_k^2}(\alpha) &= \frac{2}{J^2} \sum_{|2l| \leq n-1} \delta_l^2 l^2 \Re\left(\tilde{c}_{k,l} \sum_{j \neq k} \bar{\tilde{c}}_{j,l}\right), \\ \frac{\partial^2 W_n(\alpha)}{\partial \theta_k \partial a_m} &= \frac{-2}{J^2 a_m} \sum_{|2l| \leq n-1} \delta_l^2 l \Im(\tilde{c}_{k,l} \bar{\tilde{c}}_{m,l}), & \frac{\partial^2 W_n(\alpha)}{\partial \theta_k \partial a_k} &= \frac{-2}{J^2 a_k} \sum_{|2l| \leq n-1} \delta_l^2 l \Im(\tilde{c}_{k,l} \bar{\tilde{c}}_l), \\ \frac{\partial^2 W_n(\alpha)}{\partial a_k \partial a_m} &= \frac{-2}{J^2 a_k a_m} \sum_{|2l| \leq n-1} \delta_l^2 \Re(\tilde{c}_{k,l} \bar{\tilde{c}}_{m,l}), & \frac{\partial^2 W_n(\alpha)}{\partial a_k^2} &= \frac{-2}{J^2 a_k^2} \sum_{|2l| \leq n-1} \delta_l^2 \Re\left(\frac{\tilde{c}_{kl}}{J} \sum_{j \neq k} \bar{\tilde{c}}_{j,l} - \frac{2J-1}{J} |\tilde{c}_{k,l}|^2\right). \end{aligned}$$

Proposition 2.3. *Sous les mêmes hypothèses et notations que le théorème 2.2, $\sqrt{n} \nabla W_n(\theta^*, a^*)$ converge en loi vers un vecteur gaussien $\mathcal{N}_{2J-2}(0, \sigma^{*2} \Gamma_0)$. La matrice de covariance Γ_0 est définie de façon suivante :*

$$\Gamma_0 = \frac{2}{J^2} \begin{pmatrix} \sum_{l \in \mathbb{Z}} \delta_l^4 l^2 |c_l(f)|^2 \mathbf{V}_0 & 0 \\ 0 & \sum_{l \in \mathbb{Z}} \delta_l^4 |c_l(f)|^2 \mathbf{D}^{-1} \mathbf{V} \mathbf{D}^{-1} \end{pmatrix},$$

avec $V_0 = \left(\sum_{j=1}^J (J a_j^{*2})^{-1} \right) \mathbb{I}_{J-1} \mathbb{I}_{J-1}^T + D_a^{-2} - (\mathbb{I}_{J-1} (D_a^{-2} \mathbb{I}_{J-1})^T + (D_a^{-2} \mathbb{I}_{J-1})^T \mathbb{I}_{J-1}) / J$.

Preuve de la proposition 2.3. D'après le lemme 2.4, et après quelques calculs, on a pour $k = 2 \dots J$:

$$\begin{aligned} \sqrt{n} \frac{\partial M_n}{\partial \theta_k}(\alpha^*) &= \frac{2}{J} \sum_{l=-(n-1)/2}^{(n-1)/2} \delta_l^2 l \{ |c_l(f)| (V_l^k - V_l) + W_l^k \}, \\ \sqrt{n} \frac{\partial M_n}{\partial a_k}(\alpha^*) &= \frac{2}{J a_k^*} \sum_{l=-(n-1)/2}^{(n-1)/2} \delta_l^2 \{ |c_l(f)| (Y_l - Y_l^k) + Z_l^k \}, \end{aligned}$$

où :

$$\begin{aligned}
W_l^k &= \frac{\sigma^{*2}}{\sqrt{n}} \sum_{j=1}^J \frac{\cos \{l(\theta_k^* - \theta_j^*)\}}{a_j^* a_k^*} (\xi_{j,l}^x \xi_{k,l}^y - \xi_{j,l}^y \xi_{k,l}^x) + \frac{\sin \{l(\theta_k^* - \theta_j^*)\}}{a_j^* a_k^*} (\xi_{j,l}^x \xi_{k,l}^x + \xi_{j,l}^y \xi_{k,l}^y), \\
V_l^k &= \frac{\sigma^*}{a_k^*} \left\{ \cos(l\theta_k^* - \gamma_l) \xi_{k,l}^y + \sin(l\theta_k^* - \gamma_l) \xi_{k,l}^x \right\}, \quad \text{et} \quad V_l = \frac{1}{J} \sum_{j=1}^J V_l^j, \\
Z_l^k &= \frac{\sigma^{*2}}{\sqrt{n}} \left[\sum_{j=1}^J \left\{ \frac{\cos l(\theta_k^* - \theta_j^*)}{J a_j^* a_k^*} (\xi_{j,l}^x \xi_{k,l}^x + \xi_{j,l}^y \xi_{k,l}^y) + \frac{\sin l(\theta_k^* - \theta_j^*)}{J a_j^* a_k^*} (\xi_{j,l}^x \xi_{k,l}^y - \xi_{j,l}^y \xi_{k,l}^x) \right\} + \frac{1}{a_k^{*2}} (\xi_{k,l}^x{}^2 + \xi_{k,l}^y{}^2) \right], \\
Y_l^k &= \frac{\sigma^*}{a_k^*} (\cos(l\theta_k^* - \gamma_l) \xi_{k,l}^x - \sin(l\theta_k^* - \gamma_l) \xi_{k,l}^y), \quad \text{et} \quad Y_l = \frac{1}{J} \sum_{j=1}^J Y_l^j.
\end{aligned}$$

Pour étudier la convergence en distribution du vecteur $n^{1/2} \nabla W_n(\alpha^*)$, on étudie la convergence en loi des combinaisons linéaires de ses coordonnées. Soit $U \in \mathbb{R}^{2(J-1)}$ un vecteur, on pose $(\mu_2, \dots, \mu_J, u_2, \dots, u_J)$ ses coordonnées. On considère le produit scalaire de $\mathbb{R}^{2(J-1)}$ entre $n^{1/2} \nabla W_n(\alpha^*)$ et U , noté $\langle n^{1/2} \nabla W_n(\alpha^*), U \rangle$. Il est la somme de deux termes R_n , linéaire par rapport aux variables $\xi_{1,l}^x, \dots, \xi_{J,l}^x, \xi_{1,l}^y, \dots, \xi_{J,l}^y$, et S_n , quadratique par rapport aux variables $\xi_{1,l}^x, \dots, \xi_{J,l}^x, \xi_{1,l}^y, \dots, \xi_{J,l}^y$:

$$\begin{aligned}
\langle \sqrt{n} \nabla W_n(\alpha^*), U \rangle &= R_n + S_n, \\
\text{où :} \quad R_n &= \frac{2}{J} \sum_{j=2}^J \left[\mu_j \sum_{2|l| < n} \delta_l^2 l |c_l(f)| (V_l^j - V_l) + \frac{u_j}{J a_j^*} \sum_{2|l| < n} \delta_l^2 |c_l(f)| (Y_l - Y_l^j) \right] \\
S_n &= \frac{2}{J} \sum_{j=2}^J \left[\mu_j \sum_{2|l| < n} \delta_l^2 l W_l^j + \frac{u_j}{J a_j^*} \sum_{2|l| < n} \delta_l^2 Z_l^j \right]
\end{aligned}$$

De la même façon que pour le terme Q_n/n au lemme 2.2, l'hypothèse 2.20 assure que S_n converge en probabilité vers 0.

Il ne reste plus qu'à montrer que R_n converge en loi vers une Gaussienne. Posons :

$$(X_l) = (\xi_{1,l}^x, \dots, \xi_{J,l}^x, \xi_{1,l}^y, \dots, \xi_{J,l}^y)^T, \quad \text{et} \quad A = \text{diag}(1, 1/a_2^*, \dots, 1/a_J^*, 1, 1/a_2^*, \dots, 1/a_J^*).$$

A est une matrice carrée de taille $2J$, X_l est un vecteur aléatoire de \mathbb{R}^{2J} . On définit les vecteurs f_l^k, e_l^k de \mathbb{R}^{2J} suivants :

$$(f_l^k)_j = \begin{cases} \sin(l\theta_k^* - \gamma_l) & \text{si } j = k \\ \cos(l\theta_k^* - \gamma_l) & \text{si } j = k + J \\ 0 & \text{sinon.} \end{cases}, \quad \text{et} \quad (e_l^k)_j = \begin{cases} \cos(l\theta_k^* - \gamma_l) & \text{si } j = k \\ -\sin(l\theta_k^* - \gamma_l) & \text{si } j = k + J \\ 0 & \text{sinon.} \end{cases}.$$

Ainsi les variables V_l^k et Y_l^k peuvent s'exprimer comme le produit scalaire dans \mathbb{R}^{2J} entre ces vecteurs et la variable aléatoire X_l :

$$V_l^k = \sigma^* \langle A f_l^k, X_l \rangle, \quad \text{et} \quad Y_l^k = \sigma^* \langle A e_l^k, X_l \rangle.$$

Ainsi le terme R_n se récrit comme :

$$R_n = \frac{2\sigma^*}{J} \sum_{l=-(n-1)/2}^{(n-1)/2} \delta_l^2 |c_l(f)| < g_l(U), X_l > , \text{ où } g_l(U) = g_l^\theta(\mu) + g_l^a(u),$$

avec $g_l^\theta(\mu) = l \left(\frac{-1}{J} \sum_{k=2}^J \mu_k A f_l^1 + \sum_{j=2..J} \left(\mu_j - \frac{1}{J} \sum_{k=2}^J \mu_k \right) A f_l^j \right),$

et $g_l^a(u) = -\frac{1}{J} \sum_{k=2..J} \frac{u_k}{a_k^*} A e_l^1 + \sum_{j=2..J} \left(\frac{u_j}{a_j^*} - \frac{1}{J} \sum_{k=2..J} \frac{u_k}{a_k^*} \right) A e_l^j.$

Il est alors facile de voir que

$$\text{var} \left(\frac{2}{J} \delta_l^2 |c_l(f)| < g_l(U), X_l > \right) = \frac{2}{J^2} \delta_l^4 |c_l(f)|^2 \|g_l(U)\|_2^2,$$

où $\|g_l(U)\|_2^2 = l^2 \mu^T V \mu + u^T D_a V D_a u$, $\mu = (\mu_2, \dots, \mu_J)^T$ et $u = (u_2, \dots, u_J)^T$. Par le Lemme 2.1, et les propriétés de symétrie sur les variables complexes (2.35) et (2.15), on conclut que R_n converge en loi vers

$$\mathcal{N} \left(0, \frac{2}{J^2} \left(\sum_{l \in \mathbb{Z}} \delta_l^4 l^2 |c_l(f)|^2 \right) \mu^T \mathbf{V} \mu + \frac{2}{J^2} \left(\sum_{l \in \mathbb{Z}} \delta_l^4 |c_l(f)|^2 \right) u^T \mathbf{D}_a \mathbf{V} D_a u \right).$$

□

Proposition 2.4. *Sous les hypothèses et les notations du théorème 2.2, la matrice hessienne du critère W_n converge uniformément vers la matrice inversible H au voisinage de (θ^*, a^*) , $\mathcal{A}_{1,n}^{\text{loc}} = \{\beta \in \mathcal{A}_1, \|\beta - (\theta^*, a^*)\| \leq \|(\hat{\theta}_n - \theta^*, \hat{a}_n - a^*)\|\}$,*

$$\sup_{\beta \in \mathcal{A}_{1,n}^{\text{loc}}} \|\nabla^2 M_n(\beta) - H\| = o_{\mathbb{P}}(1).$$

H est la matrice carrée de taille $2J - 2$ suivante,

$$H = \frac{2}{J^2} \begin{pmatrix} \left(\sum_{l \in \mathbb{Z}} \delta_l^2 l^2 |c_l(f)|^2 \right) P & 0 \\ 0 & \left(\sum_{l \in \mathbb{Z}} \delta_l^2 |c_l(f)|^2 \right) D_a^{-1} P D_a^{-1} \end{pmatrix}$$

avec $P = J I_{J-1} - \mathbb{I}_{J-1} \iota(\mathbb{I}_{J-1})$.

Preuve de la proposition 2.4. D'après le lemme 2.4, on calcule de façon explicite les dérivées secondes partielles pour tout $\beta = (\theta, a) \in \mathcal{A}_1$. Par exemple, la dérivée partielle

d'ordre 2 selon les variables θ_k, θ_m avec $k \neq m$ est,

$$\frac{2}{J^2} \frac{\partial^2 M_n}{\partial \theta_k \partial \theta_m}(\beta) = \frac{-a_k^* a_m^*}{a_k a_m} \sum_{|2l| \leq n-1} \delta_l^2 l^2 |c_l(f)|^2 \cos l(\theta_k - \theta_k^* + \theta_m^* - \theta_m) \quad (2.39)$$

$$- \frac{a_k^*}{\sqrt{n} a_k a_m} \sum_{|2l| \leq n-1} \delta_l^2 l^2 |c_l(f)| \left[\cos\{l(\theta_k - \theta_k^* - \theta_m) + \gamma_l\} \xi_{k,l}^x + \sin\{l(\theta_k - \theta_k^* - \theta_m) + \gamma_l\} \xi_{k,l}^y \right] \quad (2.40)$$

$$- \frac{a_m^*}{\sqrt{n} a_k a_m} \sum_{|2l| \leq n-1} \delta_l^2 l^2 |c_l(f)| \left[\cos\{l(\theta_m - \theta_m^* - \theta_k) + \gamma_l\} \xi_{m,l}^x + \sin\{l(\theta_m - \theta_m^* - \theta_k) + \gamma_l\} \xi_{m,l}^y \right] \quad (2.41)$$

$$- \frac{1}{n a_k a_m} \sum_{|2l| \leq n-1} \delta_l^2 l^2 \left[\cos l(\theta_k - \theta_m) (\xi_{k,l}^x \xi_{m,l}^x + \xi_{k,l}^y \xi_{m,l}^y) + \sin l(\theta_k - \theta_m) (\xi_{k,l}^y \xi_{m,l}^x - \xi_{k,l}^x \xi_{m,l}^y) \right]. \quad (2.42)$$

Comme on peut l'observer, les dérivées partielles sont la somme de trois termes : un terme déterministe comme (2.39), un terme stochastique de magnitude $n^{-1/2}$ comme (2.40) et (2.41), et un terme stochastique de magnitude n^{-1} comme (2.42).

En reprenant la preuve du lemme (2.2), on prouve que les hypothèses permettent d'établir l'uniforme convergence en probabilité de la matrice hessienne de W_n vers la matrice hessienne de la fonction de contraste K sur un voisinage compact de (θ^*, a^*) .

D'autre part, en utilisant la convergence en probabilité de l'estimateur $(\hat{\theta}_n, \hat{a}_n)$ et l'uniforme continuité de $\nabla^2 K$ sur un voisinage compact de (θ^*, a^*) , la proposition est prouvée (l'argumentation détaillée de ce dernier point peut être trouvée dans la preuve de la proposition 2.6). \square

Lemmes techniques pour la preuve du théorème 2.3

Les notations qui suivent sont utilisées tout au long des preuves des lemmes et des propositions suivantes.

Avec les contraintes d'identifiabilité posées sur le modèle, et par l'orthogonalité de la base de Fourier discrète (2.9), les coefficients (\hat{c}_l) (2.25) qui maximisent le vraisemblance peuvent se récrire de la façon suivante :

$$\begin{aligned} \hat{c}_l(\alpha) &= \frac{1}{n^d J} \sum_{j=1}^J \sum_{i \in \mathcal{I}_n} a_j Y_{ij} e^{il \cdot (t_i - \theta_j)} \quad 1 \leq |l|_\infty \leq m_n, \\ &= \frac{1}{J} \sum_{j=1}^J a_j d_{jl} e^{il \cdot \theta_j} \quad 1 \leq |l|_\infty \leq m_n. \end{aligned}$$

Chacun de ces coefficients est la somme de deux termes, l'un est déterministe, l'autre est stochastique. On décrit séparément ces deux termes.

Supposons que la série de Fourier de f^* soit normalement convergente. Soit $l \in \mathbb{Z}^d$ tel que $|l|_\infty \leq m_n$. D'après le lemme 2.5, le coefficient de Fourier discret $c_l^n(f^*)$ peut se récrire sous la forme suivante :

$$c_l^n(f) = c_l(f) + \sum_{\substack{2|p|_\infty \geq n \\ p-l \in n\mathbb{Z}^d}} c_p(f).$$

Ainsi la partie déterministe des coefficients (\hat{c}_l) est égale à :

$$\frac{1}{J} \sum_{j=1}^J a_j c_l^n(f_j^*) e^{il \cdot \theta_j} = c_l(f^*) \phi(l \cdot \theta - l \cdot \theta^*, a) + g_n^l(\alpha), \quad (2.43)$$

$$\text{avec} \quad g_n^l(\alpha) = \sum_{\substack{2|p|_\infty \geq n \\ p-l \in n\mathbb{Z}^d}} c_p(f^*) \phi(l \cdot \theta - p \cdot \theta^*, a),$$

$$\phi : (x, a) \in \mathbb{R}^J \times \mathbb{R}^J \longrightarrow \sum_{j=1}^J a_j a_j^* e^{ix_j} / J.$$

Dans la suite, chaque vecteur $\theta = (\theta_{1,1}, \dots, \theta_{1,q}, \theta_{2,1}, \dots, \theta_{J,q}) \in \mathbb{R}^{Jq}$ induit la définition de J vecteurs de \mathbb{R}^d , notés $\theta_1, \dots, \theta_J$, suivants :

$$\theta_j = (\theta_{j,1}, \dots, \theta_{j,q})^T \in \mathbb{R}^d, \quad j = 1, \dots, J.$$

Pour tout $l \in \mathbb{Z}^d$, on note alors par $l \cdot \theta$ le vecteur de \mathbb{R}^J suivant $(l \cdot \theta_1, \dots, l \cdot \theta_J)$ (où $l \cdot \theta_1$ est le produit scalaire standard de \mathbb{R}^d).

Concernant la partie stochastique, les notations introduites dans la sous-section 2.2.1 sont reprises,

$$\text{pour } j = 1 \dots J, \quad \text{et } 1 \leq |l|_\infty \leq m_n \quad w_{j,l} = \xi_{j,l} / n^{d/2}.$$

La partie stochastique des coefficients \hat{c}_l est alors égale à $\sigma^* \xi_l(\alpha) / n^{d/2}$, où

$$\xi_l(\alpha) = \frac{1}{J} \sum_{j=1}^J a_j e^{il \cdot \theta_j} \xi_{j,l} \quad \text{et} \quad |\xi_l(\alpha)| \leq \frac{\sigma^*}{J\sqrt{n}} \sum_{j=1}^J |\xi_{j,l}|. \quad (2.44)$$

Lemme 2.5. Soit $l \in \mathbb{Z}^d$ tel que $1 \leq |l|_\infty \leq m_n$. Soit $p \in \mathbb{Z}^d$ tel que $|p|_\infty > m_n$ et $l - p \in n\mathbb{Z}^d$. Alors on a,

$$2|p|_\infty \geq n.$$

Preuve du lemme 2.5. Soit $l = (l_1, \dots, l_d) \in \mathbb{Z}^d$ tel que $|l|_\infty \leq m_n$. Soit $p = (p_1, \dots, p_d) \in \mathbb{Z}^d \setminus \{l\}$ tel qu'il existe $k = (k_1, \dots, k_d) \in \mathbb{Z}^d$ et $p = l + nk$. Supposons par l'absurde que $m_n < |p|_\infty < n/2$. Par définition il existe $r \in \{1 \dots d\}$ tel que $|p_r| = |p|_\infty$. Quitte à considérer $-p$, on peut supposer que p_r est positif.

En combinant l'inégalité $m_n < p_r < n/2$ et l'égalité $p - r = l_r + nk_r$, nous avons que,

$$m_n - k_r n < l_r < n/2 - k_r n.$$

Or par hypothèse $|l_r| \leq m_n$, donc k_r est nécessairement positif ($k_r \geq 1$). Par conséquent, l'inégalité suivante est vérifiée,

$$n/2 - k_r n \leq -n/2.$$

Puisque $0 \leq m_n < n/2$ et $-m_n \leq l_r$, ceci est impossible. \square

Les deux lemmes suivants sont relatifs à la preuve de la consistance de l'estimateur défini comme le minimum des moindres carrés profilés L_n (2.26).

Lemme 2.6. *Sous les hypothèses du théorème 2.3, on a :*

$$\sup_{\alpha \in \mathcal{A}_3} |L_n(\alpha) - M(\alpha) - \sigma^{*2}| = o_{\mathbb{P}}(1),$$

où par l'égalité de Parseval la fonction de contraste est la somme des deux fonctions suivantes,

$$M^1(\alpha) = \sum_{l \in \mathbb{Z}^*} |c_l(f)|^2 (1 - |\phi(l \cdot \theta - l \cdot \theta^*, a)|^2) \quad \text{et} \quad M^2(\alpha) = \frac{1}{J} \sum_{j=1}^J (v_j^* - v_j)^2.$$

Preuve du lemme 2.6. Tout d'abord, on récrit le processus de contraste L_n comme la somme de trois termes,

$$L_n(\alpha) = D_n(\alpha) + \sigma^* G_n(\alpha) + \sigma^{*2} Q_n(\alpha),$$

où

– $D_n(\alpha) = D_n^1(\alpha) - D_n^2(\alpha)$ est la partie déterministe avec,

$$D_n^1(\alpha) = \frac{1}{Jn^d} \sum_{j=1}^J \left\{ \sum_{i \in \mathcal{I}_n} a_j^* f^*(t_i - \theta_j^*) + v_j^* - v_j \right\},$$

$$D_n^2(\alpha) = \sum_{1 \leq |l|_{\infty} \leq m_n} \left| \sum_{p \in l + n\mathbb{Z}^d} c_p(f^*) \phi(l \cdot \theta - p \cdot \theta^*, a) \right|^2.$$

– $G_n(\alpha) = G_n^1(\alpha) - G_n^2(\alpha)$ est la partie stochastique qui dépend linéairement du bruit,

$$G_n^1(\alpha) = \frac{2\sigma^*}{n^d J} \sum_{j=1}^J \sum_{i \in \mathcal{I}_n} (a_j^* f^*(t_i - \theta_j^*) + v_j^* - v_j) \epsilon_{i,j},$$

$$G_n^2(\alpha) = \frac{2}{\sqrt{n}^d} \sum_{1 \leq |l|_{\infty} \leq m_n} \Re \left\{ \sum_{p \in l + \mathbb{Z}} c_p(f^*) \phi(l\theta - p\theta^*, a) \overline{\xi_l(\alpha)} \right\}.$$

– $Q_n(\alpha) = Q_n^1(\alpha) - Q_n^2(\alpha)$ est la partie stochastique quadratique en fonction du bruit,

$$Q_n^1(\alpha) = \frac{1}{n^d J} \sum_{j=1}^J \sum_{i \in \mathcal{I}_n}^n \epsilon_{i,j}^2, \quad \text{et} \quad Q_n^2(\alpha) = \frac{1}{n^d} \sum_{1 \leq |l|_\infty < m_n} |\xi_l(\alpha)|^2.$$

D'abord on montre que $Q_n(\alpha)$ converge uniformément en probabilité vers 1. En effet, Q_n^1 ne dépend pas du paramètre α , et converge en probabilité vers 1 d'après la loi forte des grands nombres. D'autre part, de l'inégalité de Cauchy-Schwarz, on déduit que tout $\alpha \in \mathcal{A}_3$,

$$0 \leq Q_n^2(\alpha) \leq Q_n^B,$$

avec $n^d J Q_n^B = \sum_{|l|_\infty < m_n} \sum_{j=1}^J |\xi_{jl}|^2$. Ainsi l'hypothèse (2.29) induit que Q_n^B converge vers 0 au sens \mathbb{L}^1 . Par conséquent Q_n^2 converge uniformément en probabilité vers 0.

Considérons, maintenant, la convergence uniforme en probabilité de G_n vers 0.

Puisque f^* est continue sur \mathcal{D} et que pour tout $j = 1 \dots J$ $|v_j| \leq v_{max}$, on peut majorer uniformément G_n^1 par :

$$\left| \frac{2}{n^d J} \sum_{j=1}^J \sum_{i \in \mathcal{I}_n} f_j^*(t_i) \epsilon_{ij} \right| + \frac{2v_{max}}{J} \sum_{j=1}^J \left| \frac{1}{n^d} \sum_{i \in \mathcal{I}_n} \epsilon_{ij} \right|.$$

Par l'inégalité de Markov, on montre que G_n^1 converge uniformément en probabilité vers 0. D'autre part, en utilisant les notations (2.43), on peut écrire G_n^2 comme la somme de deux variables G_n^{21} et G_n^{22} :

$$G_n^{21}(\alpha) = \frac{2}{n^{d/2}} \Re \left\{ \sum_{1 \leq |l|_\infty \leq m_n} c_l(f^*) \phi(l \cdot \theta - l \cdot \theta^*, a) \overline{\xi_l(\alpha)} \right\}$$

$$G_n^{22}(\alpha) = \frac{2}{n^{d/2}} \Re \left\{ \sum_{1 \leq |l|_\infty \leq m_n} g_n^l(\alpha) \overline{\xi_l(\alpha)} \right\}.$$

La variable G_n^{12} est uniformément bornée par la variable aléatoire positive suivante :

$$G_n^{21B} = \frac{2}{n^{d/2} J} \sum_{1 \leq |l|_\infty \leq m_n} |c_l(f^*)| \sum_{j=1}^J |\xi_{j,l}|.$$

Puisque la série de Fourier de f^* converge normalement (2.28), G_n^{21B} converge en probabilité vers zéro. Donc G_n^{12} converge uniformément en probabilité vers 0.

De même, G_n^{22} est bornée uniformément par

$$G_n^{22B} = \frac{1}{n^{d/2}} \left(\sum_{2|p|_\infty > n} |c_p(f^*)| \right) \sum_{|l|_\infty \leq m_n} \sum_{j=1}^J |\xi_{j,l}|.$$

D'après les hypothèses (2.28) et (2.29), on déduit que G_n^{22B} converge vers zéro en probabilité. Et donc G_n^{22} converge uniformément en probabilité vers zéro.

On termine la preuve en montrant que D_n converge uniformément vers M . Remarquons d'abord que D_n^1 converge simplement vers D^1 en tant que somme de Riemann, et que D_n^2 converge vers D^2 , avec D^1 et D^2 définies de façon suivante

$$D^1(\alpha) = \frac{1}{J} \sum_{j=1}^J \int_0^{2\pi} (f_j^*(t) - \nu_j)^2 \frac{dt}{2\pi} \quad \text{et} \quad D^2(\alpha) = \sum_{l \in \mathbb{Z}^*} |c_l(f^*) \phi(l \cdot \theta - l \cdot \theta^*, a)|^2.$$

Ainsi D_n converge simplement vers $D^1 - D^2 = M$. Montrons que cette convergence est uniforme. D'abord, comme f^* est une fonction continue sur \mathcal{D} , on a pour tout $\alpha \in \mathcal{A}_3$,

$$|D_n^1 - D^1|(\alpha) \leq \frac{1}{J} \sum_{j=1}^J \left\{ \left| \int_{\mathcal{D}} f_j^*(t)^2 \frac{dt}{2\pi} - \frac{1}{n^d} \sum_{i \in \mathcal{I}_n} f^*(t_i)^2 \right| + 2\nu_{\max} \left| c_0(f_j^*) - \frac{1}{n^d} \sum_{i \in \mathcal{I}_n} f^*(t_i) \right| \right\}$$

Par la convergence des sommes de Riemann, D_n^1 converge uniformément vers D^1 . Concernant D_n^2 , on a pour tout $\alpha \in \mathcal{A}_3$,

$$|D_n^2 - D^2|(\alpha) \leq \sum_{|l|_\infty > m_n} |c_l(f^*)|^2 + |D_n^{2B}(\alpha)|, \\ \text{avec} \quad D_n^{2B} = 2 \sum_{1 \leq |l|_\infty \leq m_n} \Re \left\{ c_l(f^*) \phi(l \cdot \theta - l \cdot \theta^*, a) \overline{g_n^l(\alpha)} \right\} + \sum_{1 \leq |l|_\infty \leq m_n} |g_n^l(\alpha)|^2.$$

Par définition des termes (g_n^l) (2.43), D_n^{2B} est majoré par

$$|D_n^{2B}(\alpha)| \leq 2 \sum_{|l|_\infty < m} |c_l(f^*)| \sum_{|p|_\infty > m_n} |c_p(f^*)| + (2m_n)^d \left| \sum_{|p|_\infty > m_n} |c_p(f^*)| \right|^2.$$

Ainsi les hypothèses (2.23) et (2.28) assurent la convergence uniforme de D_n^2 . □

Lemme 2.7 (Unicité du minimum).

Sous les hypothèses du théorème 2.3, la fonction de contraste M admet un unique minimum en α^ .*

Démonstration. Les fonctions M^2 et M^1 (définies dans le lemme 2.6) sont des fonctions positives. Remarquons que $M(\alpha^*) = 0$. Ainsi α est un minimum de M si, et seulement si $M^1(\alpha) = 0$ et $M^2(\alpha) = 0$.

Par définition, $M^2(\alpha) = 0$ implique que pour tout $j = 1 \dots J$, $v_j = v_j^*$.

Par ailleurs, par l'inégalité de Cauchy-Schwartz, on a :

$$\forall l \in \mathbb{Z}^d, \quad \forall \alpha = (\theta, a, v) \in \mathcal{A}_3, \quad |\phi(l \cdot \theta, a)| \leq 1.$$

Par conséquent $M^2(\alpha) = 0$ si, et seulement si pour tous les coefficients de Fourier ($c_l(f^*)$) non nuls, les vecteurs $(a_j^*)_{j=1 \dots J}$ et $(a_j e^{il \cdot (\theta_j - \theta_j^*)})_{j=1 \dots J}$ sont colinéaires (cas d'égalité de l'inégalité de Cauchy-Schwarz). C'est-à-dire que pour tout $l \in \mathbb{Z}^d$ tel que $c_l(f^*) \neq 0$, il existe un scalaire complexe λ_l tel que :

$$\forall j = 1 \dots J, \quad a_j^* = \lambda_l a_j e^{il \cdot (\theta_j - \theta_j^*)}.$$

Or la contrainte d'identifiabilité $\theta_1 - \theta_1^* = 0$ implique que la suite (λ_l) est réelle et que

$$l \cdot (\theta_j - \theta_j^*) \equiv 0 \pmod{2\pi}.$$

Comme au moins l'une des hypothèses **(F1)** ou **(F2)** est vérifiée, par exemple supposons que cela soit **(F2)**, il existe pour un j fixé deux systèmes d'équations de congruences

$$\left\{ \begin{array}{l} l_{\cdot r} \cdot (\theta_j - \theta_j^*) \equiv 0 \pmod{2\pi}, \\ r = 1, \dots, d \end{array} \right. \quad \text{et} \quad \left\{ \begin{array}{l} p_{\cdot r} \cdot (\theta_j - \theta_j^*) \equiv 0 \pmod{2\pi}, \\ r = 1, \dots, d \end{array} \right.$$

On inverse le système à l'aide de la matrice des cofacteurs, on déduit que :

$$\det((l_{sr})_{s,r}) \theta_j^* \equiv 0 \pmod{2\pi\mathbb{Z}^d}, \quad \text{et} \quad \det((p_{sr})_{s,r}) \theta_j^* \equiv 0 \pmod{2\pi\mathbb{Z}^d},$$

où \det est la fonction déterminant des matrices carrées. En utilisant le fait que $\det((l_{sr})_{s,r})$ et $\det((p_{sr})_{s,r})$ sont des entiers premiers entre eux, on déduit que $\theta_j = \theta_j^*$.

Par ailleurs, les contraintes d'identifiabilité $\sum_{j=1}^J a_j^2 = J$ et $a_1 > 0$ contraignent que pour tout $l \in \mathbb{Z}^d$ tel que $c_l(f^*) \neq 0$, λ_l soit égal à 1. Donc pour tout $j = 1 \dots J$, $a_j = a_j^*$.

En conclusion, si α est un minimum de M , alors $\alpha = \alpha^*$.

□

2.6 Preuves de la Section 2.3

Preuve de la proposition 2.1 La preuve de cette proposition s'effectue en deux étapes. Dans un premier temps, on prouve que la forme bilinéaire $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ définit un produit scalaire sur \mathcal{H} (cette étape est démontrée au lemme 2.8). Puis, dans un deuxième temps, on démontre le développement asymptotique (2.31).

Soit $h = (\theta_2 \dots \theta_J, a_2 \dots a_J, v_1 \dots v_J, \sigma, f)$ un élément de \mathcal{H} . Afin de satisfaire les contraintes d'identifiabilité associées au modèle (voir jeu n°3 dans la sous-section 2.1.2), on définit la suite de paramètre $(\alpha_n(h), \sigma_n(h), f_n(h))$ comme égale à :

$$\begin{aligned}\theta_j^{(n)}(h) &= \theta_j^* + \frac{1}{n^{d/2}}(\theta_j), \quad j = 2 \dots J, \\ v_j^{(n)}(h) &= v_j^* + \frac{1}{n^{d/2}}v_j \quad j = 1 \dots J, \\ a_j^{(n)}(h) &= a_j^* + \frac{1}{n^{d/2}}(a_j), \quad j = 2 \dots J, \\ a_1^{(n)}(h) &= \sqrt{J - \sum_{j=2}^J \left(a_j^* + \frac{a_j}{n^{d/2}}\right)^2}, \\ f_n(h) &= f_n = f^* + \frac{1}{n^{d/2}}f, \quad \text{et} \quad \sigma_n(h) = \sigma^* + \frac{\sigma}{n^{d/2}}.\end{aligned}$$

Comme f , f^* et ∇f^* sont des fonctions uniformément continues, on a par un développement de Taylor pour $j = 2 \dots J$ et $i \in \mathcal{I}_n$,

$$\begin{aligned}f_n(t_i - \theta_n^{(j)}(h)) - f_n(t_i - \theta_j^*) &= \frac{1}{n^{d/2}} \nabla f^*(t_i - \theta_j^*) \cdot \theta_j + o(1/n^{d/2}) \\ (a_n^{(1)}(h) - a_1^*) f_n^*(t_i) &= - \frac{\sum_{j=2}^J a_j a_j^*}{a_1^* n^{d/2}} f^*(t_i) + o(1/n^{d/2}) \\ \log \left(1 + \frac{\sigma/\sigma^*}{n^{d/2}} \right) &= \frac{\sigma/\sigma^*}{n^{d/2}} - \frac{(\sigma/\sigma^*)^2}{n} + o(n^{-d}).\end{aligned}$$

Ainsi, avec les notations de la proposition, on a

$$\Lambda_n(\alpha_n(h), \sigma_n(h), f_n(h)) = \Delta_n(h) - \frac{1}{2n^d} \sum_{i \in \mathcal{I}_n} \sum_{j=1}^J A_{ij}^n(h)^2 - \frac{J}{2} \frac{\sigma^2}{\sigma^{*2}} + o_{\mathbb{P}}(1).$$

$\sum_{i \in \mathcal{I}_n} \sum_{j=1}^J A_{ij}^n(h)^2 / n^d$ est une somme de Riemann, elle converge vers $\|h\|_{\mathcal{H}}^2$. De plus, d'après le théorème limite central de Lindeberg-Feller (voir [Van der Vaart, 1998](#), chap. 2), $\Delta_n(h)$ converge en loi vers $\mathcal{N}(0, \|h\|_{\mathcal{H}}^2)$. \square

Lemme 2.8. *Sous les hypothèses de la proposition 2.1, $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ est un produit scalaire dans \mathcal{H} .*

Preuve du lemme 2.8. Par définition, $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ est une forme bilinéaire symétrique et positive. Il reste donc à vérifier que $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ est définie, c'est-à-dire que si $h \in \mathcal{H}$ est tel que $\|h\|_{\mathcal{H}} = 0$, alors h est l'élément nul de \mathcal{H} .

Soit $h = (\alpha, \sigma, f) \in \mathcal{H}$ tel que $\|h\|_{\mathcal{H}} = 0$. Par définition de la forme $\langle \cdot, \cdot \rangle_{\mathcal{H}}$, σ est nécessairement égal à zéro, ainsi que les réels v_j $j = 1, \dots, J$. De plus, on a pour

$j = 2, \dots, J$, que :

$$\|a_j^* f + a_j f^* - a_j^* \nabla f^* \cdot \theta_j\|_{\mathbb{L}^2(\mathcal{D})} = 0 \quad \text{et} \quad \|a_1^* f - \frac{\rho}{a_1^*} f^*\|_{\mathbb{L}^2(\mathcal{D})} = 0, \quad (2.45)$$

où $\rho = \sum_{k=2}^J a_k a_k^*$. Les fonctions f , f^* et ∇f^* étant continues, et les fonctions f^* et ∇f^* étant orthogonales dans l'espace $\mathbb{L}^2(\mathcal{D})$, les équations (2.45) impliquent que la fonction f est proportionnelle à f^* ($a_1^{*2} f = \rho f^*$) et donc que,

$$\left\| \left(\frac{a_j^* \rho}{a_1^{*2}} + a_j \right) f^* \right\|_{\mathbb{L}^2(\mathcal{D})} = 0 \quad \text{et} \quad \|a_j^* \nabla f^* \cdot \theta_j\|_{\mathbb{L}^2(\mathcal{D})} = 0.$$

Puisque les dérivées partielles forment une famille libre dans l'espace vectoriel $\mathbb{L}^2(\mathcal{D})$, les vecteurs $\theta_j \in \mathbb{R}^d$ ($j = 2, \dots, J$) sont tous nuls.

D'autre part, la fonction f^* est non identiquement nulle. Ainsi, les équations suivantes sont vérifiées,

$$j = 2 \dots J, \quad a_j^* \rho / a_1^{*2} + a_j = 0.$$

Ceci implique que $\rho \frac{J - a_1^{*2}}{a_1^{*2}} + \rho = 0$, c'est à dire que la constante ρ est égale à zéro. La fonction f est donc identiquement nulle, et les réels a_j ($j = 2, \dots, J$) sont tous égaux à zéro.

En conclusion l'élément h est bien l'élément nul. \square

Lemme 2.9. *L'espace vectoriel \mathcal{H} est un sous-espace vectoriel de l'espace de Hilbert*

$$\text{cl}(\mathcal{H}) = \mathbb{R}^{J-1} \times \mathbb{R}^{J-1} \times \mathbb{R}^J \times \mathbb{R} \times \{f \in \mathbb{L}^2(\mathcal{D}), \quad c_0(f) = 0\},$$

muni du produit scalaire $\langle \cdot, \cdot \rangle_{\mathcal{H}}$.

Preuve du lemme 2.9. Il suffit de montrer que $\text{cl}(\mathcal{H})$ est un espace complet muni de la norme $\|\cdot\|_{\mathcal{H}}$. Soit $h_n = (\alpha_n, \sigma_n, f_n)$ une suite de Cauchy de $\text{cl}(\mathcal{H})$. Plus précisément, on note

$$\alpha_n = (\theta_2^{(n)}, \dots, \theta_J^{(n)}, a_2^{(n)}, \dots, a_J^{(n)}, v_1^{(n)}, \dots, v_J^{(n)}).$$

Par définition de $\|\cdot\|_{\mathcal{H}}$, les suites $(\sigma_n)_n, (v_1^{(n)})_n, \dots, (v_J^{(n)})_n$ sont des suites de Cauchy dans \mathbb{R} . Comme \mathbb{R} est complet, il existe donc $\sigma \in \mathbb{R}_+$, et $(v_1, \dots, v_J) \in \mathbb{R}^J$ tels que $(\sigma_n)_n, (v_1^{(n)}, \dots, v_J^{(n)})_n$ convergent respectivement vers σ et (v_1, \dots, v_J) .

Par ailleurs, les suites de fonctions suivantes,

$$\begin{aligned} g_1^{(n)} &= a_1^* f_n - \frac{\rho_n}{a_1^*} f^* \quad \text{où} \quad \rho_n = \sum_{k=2}^J a_k^{(n)} a_k^* \\ g_j^{(n)} &= a_j^* f_n + a_j^{(n)} f^* - a_j^* \nabla f^* \cdot \theta_j^{(n)}, \quad \text{pour } j = 2 \dots J. \end{aligned}$$

sont des suites de Cauchy dans $\mathbb{L}^2(\mathcal{D})$, ainsi que leurs combinaisons linéaires $g_j^{(n)} - g_1^{(n)} \frac{a_j^*}{a_1^*}$, $j = 2, \dots, J$.

Puisque les fonctions f^* et ∇f^* sont orthogonales dans $\mathbb{L}^2(\mathcal{D})$ et non nulles, les suites $(\theta_2^{(n)}), \dots, (\theta_J^{(n)})_n$ et $(a_j^{(n)})_{j=2\dots J}$ sont des suites de Cauchy respectivement de \mathbb{R}^d et \mathbb{R}^{J-1} .

Ainsi, la suite de fonctions $(f_n)_n$ est, elle aussi, une suite de Cauchy de $\mathbb{L}^2(\mathcal{D})$. Comme \mathbb{R} et $\mathbb{L}^2(\mathcal{D})$ sont des espaces complets, il existe des vecteurs $\theta_2, \dots, \theta_J$ de \mathbb{R}^d , $(a_j)_{j=2\dots J} \in \mathbb{R}^{J-1}$ et une fonction $f \in \mathbb{L}^2(\mathcal{D})$ tels que $(\theta_2^{(n)}), \dots, (\theta_J^{(n)})_n$, $(a_j^{(n)})_{j=2\dots J}$ et (f_n) convergent dans leur espaces respectifs vers $\theta_2, \dots, \theta_J$, $(a_j)_{j=2\dots J}$ et f .

Posons $h = (\theta_2, \dots, \theta_J, a_2 \dots a_J, v_1 \dots v_J, \sigma, f)$. h est bien un élément de $cl(\mathcal{H})$. De plus, il est évident que $(h_n)_n$ converge vers h dans $cl(\mathcal{H})$. \square

Lemme 2.10 (Calcul des représentants de la différentielle de ν_n). *Sous les hypothèses de la proposition 2.1, la différentielle du paramètre*

$$\nu_n\left(\mathbb{P}_{\alpha^*, \sigma^*, f^*}^{(n)}\right) = (\theta_2^*, \dots, \theta_J^*, a_2^*, \dots, a_J^*, v_1^*, \dots, v_J^*),$$

admet comme représentant dans l'espace de Hilbert $cl(\mathcal{H})$

$$\begin{aligned} \dot{\nu}_{j,q}^\theta &= \sigma^{*2} \left(\dot{\theta}^{j,q}, 0, 0, 0, (\nabla f^* \otimes V_a) \cdot \dot{\theta}^{j,q} / J \right) && \text{pour } j = 2 \dots J, q = 1 \dots d, \\ \dot{\nu}_j^a &= \frac{\sigma^{*2}}{\|f^*\|_{\mathbb{L}^2}^2} (0, \dot{a}^j, 0, 0, 0) && \text{pour } j = 2 \dots J \\ \dot{\nu}_j^v &= \sigma^* (0, 0, e_j, 0, 0) && \text{pour } j = 1 \dots J. \end{aligned}$$

Les vecteurs $\theta^{j,q} \in \mathbb{R}^{d \otimes (J-1)}$ sont solutions du système

$$\text{Gramm}(\nabla f^*) \otimes \left(\frac{V_a V_a^T}{J} - D^2 \right) (\dot{\theta}^{21} \dots \dot{\theta}^{2q}, \dot{\theta}^{31} \dots \dot{\theta}^{(J-1)q}, \dot{\theta}^{J1} \dots \dot{\theta}^{Jq}) = I_{d(J-1)},$$

avec $V_a = (a_2^{*2}, \dots, a_J^{*2})^T$. Les vecteurs $\dot{a}^j = (\dot{a}_k^j)_{j=2^J} \in \mathbb{R}^{J-1}$ sont définis de la façon suivante,

$$\dot{a}_k^j = \begin{cases} -a_2^* a_k^* / J & \text{if } k \neq j \\ 1 - a_k^{*2} / J & \text{if } k = j. \end{cases}$$

et le vecteur $e_j \in \mathbb{R}^J$ désigne le $j^{\text{ème}}$ vecteur de la base canonique de l'espace vectoriel \mathbb{R}^J .

Démonstration. Soit $(\dot{\alpha}, \dot{\sigma}, \dot{g}) \in \mathcal{H}$, où $\dot{\alpha} = (\dot{\theta}, \dot{a}, \dot{v})$. Pour $(\alpha, \sigma, f) \in \mathcal{H}$ avec $\alpha = (\theta, a, v)$, on réécrit le produit scalaire sous la forme suivante :

$$\begin{aligned} \sigma^{*2} \langle (\alpha, \sigma, f), (\dot{\alpha}, \dot{\sigma}, \dot{g}) \rangle &= J\sigma\dot{\sigma} + \left\langle f, J\dot{g} - \nabla f^* \cdot \dot{\lambda} \right\rangle \\ &\quad + \sum_{k=2}^J \sum_{q=1}^d \dot{\theta}_{jq} \left\langle \partial_q f^*, -a_k^{*2} \dot{g} + a_k^{*2} \sum_{r=1}^q \dot{\theta}_{kr} \partial_r f^* \right\rangle \\ &\quad + \sum_{k=2}^J a_k \left\langle f^*, \dot{a}_k f^* + \frac{a_k^*}{a_1^{*2}} \rho f^* \right\rangle + \sum_{k=2}^J v_k \dot{v}_k, \end{aligned}$$

où $\lambda = \sum_{k=2}^J \dot{\theta}_k a_k^{*2} \in \mathbb{R}^d$, (avec $\dot{\theta}_k = (\dot{\theta}_{k1} \dots \dot{\theta}_{kd})^T$) et $\rho = \sum_{k=2}^J \dot{a}_k a_k^*$.

Soient $k \in \{2, \dots, J\}$ et $q \in \{1, \dots, d\}$. On souhaite trouver un vecteur $\dot{v}_{k,q}^\theta = (\dot{\alpha}, \dot{\sigma}, \dot{g}) \in \mathcal{H}$ solution de :

$$\langle (\alpha, \sigma, f), (\dot{\alpha}, \dot{\sigma}, \dot{g}) \rangle = \theta_{kq} \quad \forall (\alpha, \sigma, f) \in \mathcal{H}.$$

De ce système, on déduit les équations suivantes pour tout $j = 2 \dots J$ et pour tout $r = 1 \dots d$:

$$\dot{\sigma} = 0, \quad \dot{v}_j = 0, \quad \text{et} \quad \dot{g} = \nabla f^* \cdot \lambda / J, \quad (2.46)$$

$$(\dot{a}_j + \rho a_j^* / a_1^{*2}) \|f^*\|^2 = 0, \quad (2.47)$$

$$\sum_{q'=1}^d \langle \partial_{q'} f^*, \partial_r f^* \rangle_{\mathbb{L}^2(\mathcal{D})} \left(\frac{1}{J} \sum_{j=2}^J a_j^* \dot{\theta}_{jq'} - \dot{\theta}_{kr} \right) = \begin{cases} \sigma^{*2} / (a_k^*)^2 & \text{si } (j, r) = (k, q) \\ 0 & \text{si } (j, r) \neq (k, q) \end{cases} \quad (2.48)$$

Par définition de ρ , on déduit de (2.47) que \dot{a} est le vecteur nul. D'autre part, le système d'équation (2.48) peut se récrire sous la forme matricielle suivante

$$\text{Gramm}(\nabla f^*) \otimes \left(\frac{V_a V_a^T}{J} - D_a^2 \right) (\dot{\theta}) = e_{q \otimes j},$$

où $e_{q \otimes j}$ est de la base canonique de l'espace tensoriel $\mathbb{R}^{q \otimes (J-1)}$. Par conséquent, on a bien que $(\dot{\alpha}, \dot{\sigma}, \dot{g}) = \dot{v}_{k,q}^\theta$.

De la même façon on résout les équations suivantes,

$$\langle (\alpha, \sigma, f), (\dot{\alpha}, \dot{\sigma}, \dot{g}) \rangle = a_k \quad \text{et} \quad \langle (\alpha, f), (\beta, g) \rangle = v_k.$$

□

2.6.1 Preuve du théorème 2.4

On reprend les notations introduites dans la section 2.5 pour la preuve du théorème 2.3. Les propositions qui suivent nécessitent les calculs des dérivées partielles de ϕ (2.43) et ξ_l (2.44).

Lemme 2.11 (dérivées de ϕ et ξ_l). *Soient $k = 2 \dots J$, et $j = 2 \dots J$,. Soient $q = 1 \dots d$ et $q' = 1 \dots d$. Alors les dérivées partielles du premier ordre sont égales à :*

$$\begin{aligned} \frac{\partial \phi}{\partial \theta_{k,q}}(l \cdot (\theta - \theta^*), a) &= \frac{il_q}{J} a_k a_k^* e^{il \cdot (\theta_k - \theta_k^*)} & \frac{\partial \phi}{\partial a_k}(l(\theta - \theta^*), a) &= \frac{1}{J} \left(-a_k \frac{a_1^*}{a_1} + a_k^* e^{l \cdot (\theta_k - \theta_k^*)} \right). \\ \frac{\partial \xi_l}{\partial \theta_{k,q}}(l\theta, a) &= \frac{il_q}{J} a_k e^{il \cdot \theta_k} \xi_{l,k} & \frac{\partial \xi_l}{\partial a_k}(l\theta, a) &= \frac{1}{J} \left(-a_k \frac{1}{a_1} \xi_{1,k} + e^{l\theta_k} \xi_{l,k} \right). \end{aligned}$$

Les dérivées partielles d'ordre 2 sont égales à :

$$\begin{aligned} \frac{\partial^2 \phi}{\partial \theta_{k,q}^2}(l \cdot (\theta - \theta^*), a) &= \frac{-l_q^2}{J} a_k a_k^* e^{il \cdot (\theta_k - \theta_k^*)} & \frac{\partial^2 \xi_l}{\partial \theta_{k,q}^2}(l \cdot \theta, a) &= \frac{-l_q^2}{J} a_k e^{il \cdot \theta_k} \xi_{l,k} \\ \frac{\partial^2 \phi}{\partial a_k^2}(l \cdot (\theta - \theta^*), a) &= \frac{1}{J} \left(-\frac{a_1^*}{a_1} - a_k^2 \frac{a_1^*}{a_1^3} \right) & \frac{\partial^2 \xi_l}{\partial a_k^2}(l \cdot \theta, a) &= \frac{1}{J} \left(-\frac{a_1^*}{a_1} - a_k^2 \frac{a_1^*}{a_1^3} \right) \xi_{1,k}. \end{aligned}$$

Proposition 2.5. *Sous les hypothèses du théorème 2.4, on a :*

$$n^{d/2} \nabla L_n(\alpha^*) = -\frac{2\sigma^*}{J} \mathbb{G}_n + o_{\mathbb{P}}(1) \quad \text{où} \quad \mathbb{G}_n^T = \left(\mathbb{G}_n^{\theta T}, \mathbb{G}_n^{aT}, \mathbb{G}_n^{vT} \right).$$

$\mathbb{G}_n^\theta, \mathbb{G}_n^a, \mathbb{G}_n^v$ sont trois vecteurs gaussiens de dimensions respectives $d(J-1)$, $J-1$ et J :

$$\begin{aligned} \mathbb{G}_n^\theta &= \frac{1}{n^{d/2}} \sum_{i \in \mathcal{I}_n} \left(I_q \otimes \left[\frac{a_1^*}{J} V_a \mid -D_a + \frac{1}{J} V_a a^T \right] \right) \nabla F^*(t_i) \epsilon_i. \\ \mathbb{G}_n^a &= \frac{1}{n^{d/2}} \sum_{i \in \mathcal{I}_n} \left[\frac{-1}{a_1^*} a \mid I_{J-1} \right] F^*(t_i) \epsilon_i. \\ \mathbb{G}_n^v &= \frac{1}{n^{d/2}} \sum_{i \in \mathcal{I}_n} \epsilon_i. \quad \text{et} \quad t_{\epsilon_i} = {}^t(\epsilon_{i1} \dots \epsilon_{iJ}) \quad \text{for } i \in \mathcal{I}_n. \end{aligned}$$

De plus, \mathbb{G}_n converge en loi vers $\mathcal{N}_{d(J-1)+2J-1}(0, H)$.

Démonstration. Nous procédons en deux étapes. D'abord, en utilisant les notations introduites dans le lemme 2.6, on montre que,

$$\sqrt{n} \nabla L_n(\alpha^*) = \sqrt{n} (\nabla G_n^1(\alpha^*) - G_n^{21}(\alpha^*)) = -\frac{2\sigma^*}{J} \mathbb{G}_n + o_{\mathbb{P}}(1).$$

Ensuite, on vérifie que \mathbb{G}_n est un vecteur Gaussien convergeant vers $\mathcal{N}_{d(J-1)+2J-1}(0, H)$.

On étudie séparément les termes G_n , D_n et Q_n . Soit $k \in \{2 \dots J\}$ et soit $q \in \{1 \dots d\}$, alors d'après le lemme 2.11, on a

$$\frac{\partial Q_n}{\partial \theta_{k,q}}(\alpha^*) = -\frac{a_k^*}{Jn^d} \sum_{1 \leq |l|_\infty \leq m_n} \Re \left(\frac{il_q e^{il \cdot \theta_k^*}}{J} \xi_{k,l} \overline{\xi_l(\alpha^*)} \right).$$

Ainsi, on peut majorer ce terme de la façon suivante,

$$\left| n^{d/2} \frac{\partial Q_n}{\partial \theta_{k,q}}(\alpha^*) \right| \leq \frac{2}{J^2 n^{d/2}} \sum_{1 \leq |l|_\infty \leq m_n} |l_q| |\xi_{k,l}| \sum_{j=1}^J |\xi_{j,l}|.$$

Ainsi $n^{d/2} \frac{\partial Q_n}{\partial \theta_{k,q}}(\alpha^*)$ converge en probabilité vers zéro si $m_n^{d+1}/n^{d/2} = o(1)$. De la même façon, la dérivée partielle en a_k^* converge en probabilité vers zéro si $m_n^d/n^{d/2} = o(1)$. Comme la dérivée partielle en v_k^* est nulle, l'hypothèse (2.33) assure la convergence en probabilité de $n^{d/2} \nabla Q_n(\alpha^*)$ vers zéro.

Concernant la partie déterministe, après calculs (en utilisant les notations (2.43)), la dérivée partielle selon la variable $\theta_{k,q}$ est égale à :

$$\frac{\partial D_n}{\partial \theta_{k,q}}(\alpha^*) = -\frac{2a_k^{*2}}{J} \sum_{1 \leq |l|_\infty \leq m_n} \Re \left\{ i l_q c_l(f) \overline{g_n^l(\alpha^*)} + i l_q |g_n^l(\alpha^*)|^2 \right\}$$

Par définition de la suite $(g_n^l(\alpha^*))_l$ et le lemme 2.5, on déduit alors la majoration suivante :

$$n^{d/2} \left| \frac{\partial D_n}{\partial \theta_{k,q}}(\alpha^*) \right| = \frac{2a_k^{*2}}{J} n^{d/2} \left\{ 2 \sum_{|l|_\infty \leq m_n} |l_q c_l(f^*)| \sum_{2|p|_\infty \geq n} |c_p(f^*)| + \sum_{|l|_\infty \leq m_n} |l_q| \left(\sum_{2|p|_\infty \geq n} |c_p(f^*)| \right)^2 \right\}.$$

Par l'hypothèse (2.32), on a que,

$$n^{d/2} \frac{\partial D_n}{\partial \theta_{k,q}}(\alpha^*) = o \left(1 + \frac{m_n^{d+1}}{n^{d/2}} \right).$$

D'après l'hypothèse (2.33), $n^{d/2} \partial D_n / \partial \theta_{k,q}$ en α^* converge vers zéro. De la même façon, $n^{d/2} \partial D_n / \partial a_k$ en α^* converge vers 0. Concernant les dérivées partielles selon le paramètre v , on a :

$$n^{d/2} \frac{\partial D_n}{\partial v_k}(\alpha^*) = -\frac{2n^{d/2}}{J n^{d/2}} \left(\sum_{i \in \mathcal{I}_n} f_k^*(t_i) - v_k^* \right) = -\frac{2a_k^*}{J} n^{d/2} \sum_{p \in n\mathbb{Z}^d \setminus \{0\}} c_p(f^*) e^{-ip \cdot \theta_k^*}.$$

D'après l'hypothèse (2.33), $n^{d/2} \partial D_n / \partial v_k$ en α^* converge vers zéro. Par conséquent $n^{d/2} \nabla D_n(\alpha^*)$ converge vers zéro.

On vient de montrer que $n^{d/2} \nabla L_n(\alpha^*) = n^{d/2} \nabla G_n(\alpha^*) + o_{\mathbb{P}}(1)$. En prenant les notations du lemme 2.6 et après calculs, on a

$$n^{d/2} \frac{\partial G_n^{22}}{\partial \theta_{k,q}}(\alpha^*) = \mathcal{O} \left(\sum_{1 \leq |l|_\infty \leq m_n} |l_q| (|\xi_{l,j}| + |\xi_l(\alpha^*)|) \sum_{2|p|_\infty \geq n} |c_p(f^*)| \right).$$

Ainsi, d'après les hypothèses (2.32) et (2.33), $n^{d/2} \frac{G_n^{22}}{d\theta_k}(\alpha^*) = \mathcal{O}_{\mathbb{P}}(m_n^{d+1}/n^{d/2})$ converge en probabilité vers zéro. De même $n^{d/2} \frac{\partial G_n^{22}}{\partial a_k}(\alpha^*)$ converge en probabilité vers zéro. Le vecteur $n^{d/2} \nabla G_n^{22}(\alpha^*)$ converge donc en probabilité vers zéro, et on a que $n^{d/2} \nabla L_n(\alpha^*) = \sqrt{n} \nabla G_n^1(\alpha^*) - \sqrt{n} \nabla G_n^{21}(\alpha^*) + o_{\mathbb{P}}(1)$.

Plus précisément, les dérivées partielles sont égales à,

$$\begin{aligned} n^{d/2} \frac{\partial L_n}{\partial \theta_{k,q}}(\alpha^*) &= -\frac{2\sigma^*}{J} \sum_{1 \leq |l|_{\infty} \leq m_n} \Re \left\{ i l_q c_l(f^*) \left(a_k^{*2} \overline{\xi_l(\alpha^*)} - a_k^* e^{-il \cdot \theta_k^*} \overline{\xi_{k,l}} \right) \right\} + o_{\mathbb{P}}(1) \\ n^{d/2} \frac{\partial L_n}{\partial a_k}(\alpha^*) &= -\frac{2\sigma^*}{J} \sum_{1 \leq |l| \leq m_n} \Re \left\{ c_l(f^*) \left(e^{-il \cdot \theta_k^*} \overline{\xi_{k,l}} - \frac{a_k^*}{a_1^*} \overline{\xi_{1,l}} \right) \right\} + o_{\mathbb{P}}(1) \\ n^{d/2} \frac{\partial L_n}{\partial v_k}(\alpha^*) &= -\frac{2\sigma^*}{J n^{d/2}} \sum_{i \in \mathcal{I}_n} \epsilon_{i,k} = -\frac{2\sigma^*}{J} \overline{\xi_{k,0}}. \end{aligned}$$

Par définition des variables ($\xi_{kl} = n^{d/2} w_{kl}$) (voir 2.11), l'hypothèse (2.32) entraîne que pour tout $k = 1 \dots J$:

$$\begin{aligned} \Re \left\{ \sum_{|l|_{\infty} \leq m_n} i l c_l(f^*) \bar{\xi}_{kl} e^{-il \theta_k^*} \right\} &= \frac{1}{\sqrt{n}} \sum_{i \in \mathcal{I}_n} \epsilon_{ik} \Re \left\{ \sum_{|l|_{\infty} \leq m_n} i l c_l(f^*) e^{il(t_i - \theta_k^*)} \right\} \\ &= \frac{1}{\sqrt{n}} \sum_{i \in \mathcal{I}_n} \epsilon_{ik} \nabla f^*(t_i - \theta_k^*) + o_{\mathbb{P}}(1). \end{aligned}$$

On retrouve alors l'expression de \mathbb{G}_n définie dans la proposition.

D'autre part, le vecteur $(\mathbb{G}_n^{\theta}, \mathbb{G}_n^a, \mathbb{G}_n^v)$ peut se récrire sous la forme suivante,

$$\begin{aligned} \mathbb{G}_n^{\theta} &= \sum_{1 \leq |l| \leq m_n} \Re \left\{ i \left(c_l(f^*) l \otimes \left[\frac{a_1^*}{J} V_a \mid V_a \frac{a^T}{J} - D_a \right] \right) X_l^* \right\} + o_{\mathbb{P}}(1) \\ \mathbb{G}_n^a &= \sum_{1 \leq |l| \leq m_n} \Re \left\{ c_l(f^*) \left[\frac{-1}{a_1^*} a \mid I_{J-1} \right] X_l^* \right\} + o_{\mathbb{P}}(1) \quad \text{et} \quad \mathbb{G}_n^v = \Re \{ X_0^* \} + o_{\mathbb{P}}(1), \end{aligned}$$

où X_l^* est le vecteur complexe Gaussien suivant,

$$X_l^* = (e^{-il \cdot \theta_k^*} \overline{\xi_{1l}}, \dots, e^{-il \cdot \theta_k^*} \overline{\xi_{Jl}})^T.$$

Comme les variables $(X_l^*)_{|l|_{\infty} \leq m_n}$ sont indépendantes et identiquement distribuées, et comme les variables \mathbb{G}_n^{θ} et \mathbb{G}_n^a ne dépendent pas de X_0^* , \mathbb{G}_n^v est indépendant de \mathbb{G}_n^{θ} \mathbb{G}_n^a et sa matrice de covariance est la matrice identité de \mathbb{R}^J .

Par ailleurs, les partie imaginaire et réelle de $c_l(f^*) X_l^*$ sont indépendantes. Les variables \mathbb{G}_n^{θ} et \mathbb{G}_n^a sont donc asymptotiquement indépendantes. Leurs matrices de covariance sont respectivement égales à $\text{Gramm}(\nabla f^*) \otimes (D_a^2 - V_a V_a^T / J)$ et $\|f^*\|_{\mathbb{L}^2(\mathcal{D})}^2 (I_{J-1} - a a^T / a_1^{*2})$ (cf lemme 2.1). \square

Proposition 2.6. *Sous les hypothèses du théorème 2.4, la matrice hessienne du processus de contraste converge uniformément vers la matrice $2H/J$ au voisinage de α^* , $\mathcal{A}_{3,n}^{\text{loc}} = \{\alpha \in \mathcal{A}_3, \|\alpha - \alpha^*\| \leq \|\hat{\alpha}_n - \alpha^*\|\}$,*

$$\sup_{\alpha \in \mathcal{A}_{3,n}^{\text{loc}}} \|\nabla^2 L_n(\alpha) - 2H/J\| = o_{\mathbb{P}}(1).$$

Démonstration. Après calculs, utilisant le lemme 2.11, la matrice $-2H/J$ est la valeur de la matrice hessienne du critère limite M (voir la preuve du théorème 2.3) au point α^* . On étudie localement (sur l'espace $\mathcal{A}_{3,\text{loc}}$) la matrice hessienne de L_n . Remarquons aussi que pour tout $\epsilon > 0$ fixé :

$$\begin{aligned} \mathbb{P} \left(\sup_{\alpha \in \mathcal{A}_{3,\text{loc}}} \|\nabla^2 L_n(\alpha) - \nabla^2 L(\alpha^*)\| > 2\epsilon \right) &\leq \mathbb{P} \left(\sup_{\alpha \in \mathcal{A}_{3,\text{loc}}} \|\nabla^2 L_n(\alpha) - \nabla^2 M(\alpha)\| > \epsilon \right) \\ &\quad + \mathbb{P} \left(\sup_{\alpha \in \mathcal{A}_{3,\text{loc}}} \|\nabla^2 M(\alpha) - \nabla^2 M(\alpha^*)\| > \epsilon \right) \end{aligned} \quad (2.49)$$

Sur un voisinage compact de α^* , les hypothèses du théorème 2.4 assurent que $\nabla^2 L_n$ converge uniformément en probabilité vers $\nabla^2 M$. Ainsi le premier terme de la somme converge vers zéro avec n .

D'autre part, $\nabla^2 M$ est une fonction uniformément continue sur un voisinage compact de α^* . Par conséquent il existe $\eta > 0$ tel que :

$$\nabla^2 M(B(\alpha^*, \delta)) \subseteq B(\nabla^2 M(\alpha^*), \epsilon),$$

où $B(\alpha^*, \delta)$ désigne la boule ouverte de centre α^* et de rayon δ . Par conséquent l'inclusion suivante d'évènements est vérifiée,

$$\left(\sup_{\alpha \in \mathcal{A}_{3,\text{loc}}} \|\nabla^2 M(\alpha) - \nabla^2 M(\alpha^*)\| > \epsilon \right) \subseteq (\|\hat{\alpha}_n - \alpha^*\| > \delta).$$

D'après le théorème 2.3, le second terme de l'inégalité (2.49) converge vers zéro avec n . □

2.6.2 Preuves et lemmes techniques concernant l'estimateur de la forme commune

Preuve du corollaire 2.1

D'après les hypothèses du corollaire, on peut déterminer la vitesse de convergence de la série de Fourier de f^* ,

$$\sum_{|l|_{\infty} > n} |c_l(f^*)| = o(n^{-k+1/2}) \quad \text{et} \quad \sum_{|l|_{\infty} > n} |c_l(f^*)|^2 = o(n^{-2k}).$$

D'après les notations introduites dans la preuve du théorème 2.3, la différence entre l'estimateur \hat{f}_n et la fonction f^* peut s'écrire pour $t \in \mathbb{R}^d$,

$$\begin{aligned} f^*(t) - f_{\alpha^*}(t) &= \sum_{|l|_\infty > m_n} c_l(f^*) e^{il \cdot t} + \sum_{1 \leq |l|_\infty \leq m_n} e^{il \cdot t} g_n^l(\hat{\alpha}_n) \\ &+ \sum_{1 \leq |l|_\infty \leq m_n} c_l(f^*) \left\{ \phi(l(\hat{\theta} - \theta^*), \hat{a}) - 1 \right\} e^{il \cdot t} \end{aligned} \quad (2.50)$$

$$+ \sigma^* \sum_{1 \leq |l|_\infty \leq m_n} \xi_l(\hat{\alpha}) \frac{e^{il \cdot t}}{\sqrt{n}}. \quad (2.51)$$

D'abord, on étudie le terme (2.50). Par le théorème 2.4 et par la méthode Delta, on a pour tout $j = 1, \dots, J$:

$$e^{il \cdot (\hat{\theta}_j - \theta_j^*)} - 1 = il \cdot (\hat{\theta}_j - \theta_j^*) + \mathcal{O}_{\mathbb{P}}((|l_1| + \dots + |l_d|)/n^{d/2}).$$

De plus, l'inégalité suivante est vérifiée :

$$\left| \phi(l \cdot (\hat{\theta} - \theta^*), \hat{a}) - 1 \right| \leq \frac{1}{J} \sum_{j=1}^J a_j^* |\hat{a}_j - a_j^*| + \frac{1}{J} \sum_{j=1}^J a_j^{*2} |e^{il \cdot (\hat{\theta}_j - \theta_j^*)} - 1|.$$

On en déduit que,

$$\begin{aligned} \left| \phi(l \cdot (\hat{\theta} - \theta^*), \hat{a}) - 1 \right| &= ((|l_1| + \dots + |l_d|) + 1) \mathcal{O}_{\mathbb{P}}(1/n^{d/2}) \\ \mathbb{E} \left| \phi(l \cdot (\hat{\theta} - \theta^*), \hat{a}) - 1 \right|^2 &= ((l_1^2 + \dots + l_d^2) + 1) \mathcal{O}(1/n^d). \end{aligned}$$

Par ailleurs, le terme (2.51) est majoré de la façon suivante (par l'inégalité de Cauchy-Schwarz) :

$$\begin{aligned} \mathcal{Q}_n &= \left| \sum_{1 \leq |l|_\infty \leq m_n} \xi_l(\hat{\alpha}) \frac{e^{ilt}}{n^{d/2}} \right| \leq \frac{1}{J^{1/2} n^{d/2}} \sum_{1 \leq |l|_\infty \leq m_n} \sum_{j=1}^J |\xi_{jl}|, \\ &\int_0^{2\pi} \mathbb{E}(W_n^2) \frac{dt}{2\pi} = \frac{1}{n^d J} \sum_{1 \leq |l|_\infty \leq m_n} \sum_{j=1}^J |\xi_{jl}|^2. \end{aligned}$$

Par l'inégalité de Markov, on a,

$$\mathcal{Q}_n = \mathcal{O}_{\mathbb{P}}(m_n/\sqrt{n}) \quad \text{and} \quad \int_0^{2\pi} \mathbb{E}(\mathcal{Q}_n^2) \frac{dt}{2\pi} = \mathcal{O}(m_n/n).$$

Ainsi, par les hypothèses du théorème, on obtient,

$$\begin{aligned} f^*(t) - f_{\alpha^*}(t) &= \sum_{|l|_\infty > m_n} c_l(f^*) e^{ilt} + o_{\mathbb{P}}(m_n^d/n^d) + \mathcal{O}_{\mathbb{P}}(1/n^{d/2} + m_n^d/n^{d/2}) \\ \mathbb{E} \int_{\mathcal{D}} |f^*(t) - f_{\alpha^*}(t)|^2 \frac{dt}{2\pi^d} &= \sum_{|l|_\infty > m_n} |c_l(f^*)|^2 + o(m_n^d/n^d) + \mathcal{O}(1/n^d + m_n^d/n^d). \end{aligned}$$

Les hypothèses de régularité sur la fonction f^* permettent de conclure.

Propriétés asymptotiques de \hat{f}_n .

Lemme 2.12. *Soit $l \in \mathbb{Z}^d \setminus \{0\}$. Supposons que les hypothèses du théorème 2.4 soient vérifiées. Pour n assez grand, on a,*

$$\begin{aligned} n^{d/2} \Re(\hat{c}_l(\hat{\alpha}_n) - c_l(f^*)) &= \Delta_n \left(-\Im(c_l(f^*)) \tilde{h}_l^f \right) + \Delta_n(\nu_{b_{1l}^*}) + o_{\mathbb{P}}(1) \\ n^{d/2} \Im(\hat{c}_l(\hat{\alpha}_n) - c_l(f^*)) &= \Delta_n \left(\Im(l c_l(f^*)) \tilde{h}_l^f \right) + \Delta_n(-\nu_{b_{2l}^*}) + o_{\mathbb{P}}(1), \end{aligned}$$

où $\tilde{h}^f = \sigma^{*2} \left((l^T \text{Gramm}(\nabla f^*)) \otimes \left(\frac{\mathbb{I}_{J-1}}{a_1^{*2}} \right), 0, 0, 0, \frac{J-a_1^{*2}}{a_1^{*2}} (l^T \text{Gramm}(\nabla f^*)) \nabla f^* \right)$.

Démonstration. Soit $l \in \mathbb{Z}^d \setminus \{0\}$. Pour n assez grand (de telle façon que $|l|_{\infty} \leq m_n$), par le théorème de l'application continue (Van der Vaart, 1998, théorème. 2.3) et l'hypothèse (2.32) (qui assure la convergence du coefficient de Fourier discret $c_l^n(f^*)$ vers $c_l(f^*)$ à la vitesse $n^{d/2}$), nous avons que,

$$\begin{aligned} n^{d/2} c_l(\hat{f}_n - f^*) &= n^{d/2} (\hat{c}_l(\hat{\alpha}_n) - c_l(f^*)), \\ &= c_l(f^*) n^{d/2} \left(\frac{1}{J} \sum_{j=1}^J \hat{a}_j a_j^* e^{il \cdot (\hat{\theta}_j - \theta_j^*)} - 1 \right) + \xi_l(\theta^*, a^*) + o_{\mathbb{P}}(1). \end{aligned}$$

Puisque $n^{d/2}(\hat{\alpha}_n - \alpha^*)$ converge en loi (théorème 2.4), on montre en utilisant la méthode Delta (Van der Vaart, 1998, Chap. 3) que :

$$n^{d/2} c_l(\hat{f}_n - f^*) = i c_l(f^*) \left[\frac{l \otimes V_a}{J} \right] \cdot [n^{d/2}(\hat{\theta}_j - \theta_j^*)] + \xi_l(\theta^*, a^*) + o_{\mathbb{P}}(1).$$

Le théorème 2.4 et le lemme 2.10 permettent de déduire (car Δ_n est linéaire) que :

$$n^{d/2} c_l(\hat{f}_n - f^*) = i c_l(f^*) \Delta_n(\tilde{h}^f) + \xi_l(\theta^*, a^*) + o_{\mathbb{P}}(1).$$

Or par définition du processus centré Δ_n (voir la proposition 2.1) et la définition de ξ_l (2.44), on peut récrire :

$$\Re(\xi_l(\theta^*, a^*)) = \Delta_n(\nu_{b_{1l}^*}), \quad \text{et} \quad \Im(\xi_l(\theta^*, a^*)) = \Delta_n(-\nu_{b_{2l}^*}).$$

□

Chapitre 3

Test d'adéquation pour une sous-classe de modèle de forme invariante *

Sommaire

3.1	Estimation des paramètres	84
3.2	Test d'adéquation du modèle	92
3.3	Test sur l'homogénéité des données	102

Résumé : On observe un nombre fixé de fonctions inconnues au cours du temps. Les données sont modélisées sous un modèle de régression. L'objet de ce chapitre est de construire une procédure de test statistique pour tester si les fonctions de régression diffèrent des unes des autres seulement en amplitude et en phase. Deux procédures de test sont proposées. Sous l'hypothèse nulle, la loi limite de la première statistique de test étant plutôt compliquée, la première procédure de test est implémentée à l'aide d'une méthode bootstrap. Cette procédure est illustrée sur des données d'électroencéphalogramme. La seconde contourne le problème de la complexité de la loi en testant une hypothèse plus faible sur l'homogénéité des coefficients de Fourier des fonctions de régression.

Mots clés : Shape invariant model, Test d'adéquation, Estimation semi-paramétrique, Transformée de Fourier discrète.

***Remerciements :** J'aimerais remercier Hölger Dette pour ses suggestions quand à l'utilisation d'une technique Bootstrap pour implémenter le test, et Jérémie Bigot, de m'avoir transmis la base de données sur les électroencéphalogrammes.

Introduction

Dans plusieurs études statistiques, la réponse intéressante est souvent une fonction continue mesurée au cours du temps. Notons $Y_{i,j}$ l'observation du sujet j , $j = 1 \dots J$, au temps $t_{i,j}$, $i = 1 \dots n_j$. Le modèle général étudié ici est le suivant :

$$Y_{i,j} = f_j^*(t_{i,j}) + \sigma_j^* \epsilon_{i,j}, \quad i = 1 \dots n_j, \quad j = 1 \dots J. \quad (3.1)$$

C'est un modèle de régression de dépendance non linéaire, puisque les paramètres inconnus sont les fonctions f_j^* , $j = 1 \dots J$, qui dépendent non linéairement des temps d'observation $(t_{i,j})_{i,j} \in [0, T]$. L'erreur du modèle $(\epsilon_{i,j})$ sont des variables indépendantes centrées de variance égale à 1. Le réel positif σ_j^* , pour $j = 1 \dots J$ fixé est l'écart type du bruit pour le sujet j .

Dans de nombreux exemples, les formes des courbes observées semblent similaires. Il est alors raisonnable de supposer que les fonctions f_j^* diffèrent les unes des autres par une simple transformation. Prenons l'exemple de l'observation d'un phénomène biologique chez des sujets saints (température du corps, courbe de croissance, ...). Les sujets étant différents, les courbes d'évolution possèdent leurs caractéristiques propres (température moyenne du corps plus ou moins élevée, puberté plus ou moins précoce, ...), mais leurs formes sont semblables (dans le cas contraire cela signifierait qu'un des sujets est malade). L'étude exposée ici a pour but de construire une procédure statistique pour tester si effectivement les fonctions de régression ont une forme commune. Plus précisément, on souhaite tester si les fonctions f_j^* vérifient le modèle suivant :

$$f_j^*(t) = a_j^* f^*(t - \theta_j^*) + v_j^*, \quad t \in \mathbb{R}, \quad j = 1 \dots J. \quad (3.2)$$

La fonction inconnue f^* détermine la forme commune des fonctions f_j^* . Les vecteurs de \mathbb{R}^J , $\theta^* = (\theta_j^*)_{j=1 \dots J}$, $a^* = (a_j^*)_{j=1 \dots J}$ et $v^* = (v_j^*)_{j=1 \dots J}$ déterminent les caractéristiques propres à chacun des sujets. Ils sont eux aussi inconnus. Le modèle (3.2) est une sous classe du modèle de forme invariante introduit par [Lawton et al. \(1972\)](#). Des applications de ce modèle peuvent être trouvées par exemple pour l'étude du volume d'air rejeté par les poumons ([Lawton et al., 1972](#)), des rythmes biologiques ([Wang et Brown, 1996](#)), ou la prévision du trafic routier ([Loubes et al., 2006](#); [Gamboa et al., 2005](#)).

Dans ce type de modèle, c'est d'abord le problème de l'estimation du paramètre $\alpha^* = (\theta^*, a^*, v^*)$ et de la fonction f^* qui a été étudié. [Lawton et al. \(1972\)](#) définissent un algorithme itératif, SEMOR, qui utilise des splines d'ordre 1 pour estimer la forme commune f^* . La convergence et la consistance de la méthode SEMOR sont démontrées par [Kneip et Gasser \(1988\)](#). Plusieurs auteurs ont choisi d'autres bases de

fonctions pour estimer la forme commune, par exemple des splines lisses (Guardabasso et al., 1988), des splines périodiques (Wang et Brown, 1996). Une autre approche non itérative est proposée par Härdle et Marron (1990). Les auteurs étudient le cas de deux courbes qui diffèrent seulement par une translation selon les axes des abscisses et des ordonnées. Ils utilisent un estimateur à noyau pour estimer la fonction f^* et construisent ainsi un estimateur asymptotiquement normal de vitesse de convergence \sqrt{n} pour les paramètres θ^* et v^* . Plus récemment, Gamboa et al. (2005) construisent un estimateur asymptotiquement normal de vitesse de convergence \sqrt{n} pour le paramètre de translation. Cette méthode d'estimation utilise la transformée de Fourier discrète, elle est définie pour un nombre fixé $J \geq 2$ de courbes. De plus aucun estimateur de la fonction f^* est requis. Cette méthode est généralisée dans le chapitre 2 pour le modèle complet (3.2) afin d'estimer le paramètre α^* et la fonction f^* . De plus l'estimateur de α^* s'est avéré asymptotiquement efficace.

Parallèlement, le problème de l'adéquation des données au modèle a été étudié. Les tests d'adéquation permettent dans la pratique de valider l'existence de rythmes biologiques (voir Wang et Brown, 1996) ou encore de mesurer l'homogénéité des données (voir l'application avec les données d'encéphalogramme dans la sous section 3.2.3). Plusieurs procédures de test ont été définies pour tester de telles hypothèses. Möcks et al. (1984) proposent deux procédures de test visant à tester si les fonctions f_j^* étaient identiques. L'une est sensible à la variation en amplitude. L'autre teste si les projections des fonctions f_j sur des sous espaces sont différentes. D'autre part, Härdle et Marron (1990) construisent des procédures pour tester si les données vérifient le modèle de forme invariante, et si un des paramètres peut être éliminer. Leurs statistiques de test sont définies à partir de leurs critères d'estimation. Wang et Brown (1996) proposent une procédure statistique afin de tester l'ajustement d'un des paramètres au modèle.

L'objet de ce chapitre est de construire un test d'adéquation pour le modèle (3.2). On développe deux approches. La première approche (cf. les sections 3.1 et 3.2) teste effectivement si les données adhèrent au modèle (3.2). La statistique de test est définie à partir d'un processus de contraste pour l'estimation du paramètre α^* . Le théorème 3.1 affirme que cette statistique est petite lorsque le modèle (3.2) est correct, et qu'elle tend vers l'infini dans le cas contraire. Puisque la loi limite de cette statistique est plutôt compliquée sous l'hypothèse nulle, la procédure de test est implémentée par un algorithme bootstrap (cf. la section 3.2.2). Il est par conséquent nécessaire de définir des estimateurs pour les paramètres du modèle général (3.1) et du modèle sous l'hypothèse nulle (3.2) (cf. section 3.1).

La seconde approche teste une hypothèse plus faible que la précédente : l'homogé-

néité des fonctions de régression f_j^* . Pour chaque coefficient de Fourier, on définit des statistiques de test pour tester leur égalité en module. En agrégeant ces statistiques et en utilisant la théorie des valeurs extrêmes (Resnick (1987)), on définit une statistique pour tester l'homogénéité des fonctions f_j^* (théorème 3.3). Contrairement à la dernière approche, la loi limite sous H_0 est connue, c'est la loi de Gumbel.

Remarquons enfin que dans ces deux approches, les statistiques de test sont construites à partir de processus de contraste pour l'estimation du paramètre α^* similaires à ceux présentés dans le chapitre 2. Ces processus sont basés sur la transformée de Fourier des données. C'est pourquoi les fonctions de régression sont supposées périodiques. Néanmoins, comme l'explique la remarque 3.1, cette hypothèse n'est pas si restrictive.

Le plan du chapitre est le suivant. La section 3.1 présente les processus de contraste utilisés pour le test d'adéquation du modèle (3.2), ainsi que les estimateurs des paramètres pour le modèle général (3.1). Ensuite, la section 3.2 définit la procédure de test pour la bonne adéquation des données au modèle (3.2). Enfin, la section 3.3 est consacrée à la procédure de test pour l'homogénéité des fonctions de régression f_j^* .

3.1 Estimation des paramètres

Cette section présente brièvement la définition des estimateurs des paramètres pour le modèle général (3.1) et les estimateurs pour le modèle (3.2). La sous-section 3.1.1 décrit brièvement le modèle (3.2). Une description plus détaillée est faite au chapitre 2. La sous-section 3.1.2 résume les résultats du chapitre 2 pour l'estimation des paramètres sous l'hypothèse nulle. Par la suite ces estimateurs sont utilisés pour générer les échantillons bootstrap sous l'hypothèse nulle. La sous-section 3.1.3 définit les estimateurs des fonctions de régression f_j^* et de la variance pour le modèle général (3.1). Ces estimateurs sont utilisés pour normaliser la statistique de test et pour calculer les résidus dans la procédure bootstrap.

3.1.1 Description du modèle (3.2)

Les observations $(Y_{i,j})_{1 \leq i \leq n, 1 \leq j \leq J}$ proviennent de l'observation de J courbes. Sans perte de généralité et pour simplifier les calculs, on suppose que l'intervalle d'observation est $[0, 2\pi]$ et que chaque courbe est observée sur le même espace de points,

$$t_{i,j} = t_i = \frac{i-1}{n} 2\pi \in [0, 2\pi[, \quad j=1 \dots J, \quad i=1 \dots n.$$

Le choix de considérer des points d'observation équidistants est fortement conseillé du fait de l'utilisation de la transformée de Fourier discrète (voir la sous-section 2.2.1 pour plus de détails). Ainsi le modèle étudié sous l'hypothèse nulle est :

$$Y_{i,j} = a_j^* f^*(t_i - \theta_j^*) + v_j^* + \sigma^* \epsilon_{i,j}, \quad j = 1 \dots J, \quad i = 1 \dots n. \quad (3.3)$$

f^* est une fonction inconnue, 2π -périodique et continue. Les termes de bruit $(\epsilon_{i,j})_{i,j}$ sont des variables gaussiennes indépendantes, centrées et réduites. Pour simplifier, on considère le modèle où les variances sont identiques : $\sigma^{*2} = \sigma_j^{*2}$, $j = 1 \dots J$. Néanmoins, le modèle où les variances ne sont pas toutes égales est discuté tout au long de l'exposé.

Ce modèle est semi-paramétrique : $\alpha^* = (\theta^*, a^*, v^*, \sigma^*)$ est le paramètre fini-dimensionnel, et la fonction f^* est le paramètre de nuisance.

Remarque 3.1. Dans le modèle, on suppose que les fonctions de régression sont 2π -périodiques. Au premier abord, cette hypothèse peut sembler restrictive. Néanmoins, elle est vérifiée quand les fonctions sont prolongeables par T -périodicité sur \mathbb{R} . Par exemple, cette hypothèse est vraie si les fonctions ont un support compact :

$$\text{cl}\{t \in [0, T], f_j^*(t) \neq 0\} \subset (0, T) \quad \forall j = 1 \dots J,$$

où $\text{cl}(B)$ est la fermeture de l'ensemble $B \subset \mathbb{R}$.

Comme on l'a remarqué dans la sous-section 2.1.2, le modèle présenté n'est pas identifiable, il est nécessaire de fixer des contraintes d'identifiabilité. Le jeu de contrainte utilisé est ici le jeu n°3 (voir 2.1.2) :

$$f \in \mathcal{F}_0 \quad \text{et} \quad \alpha \in \mathcal{A}_3 = \{\alpha \in \mathcal{A}, \text{ tel que } \theta_1 = 0 \text{ et } \sum_{j=1}^J a_j^2 = J, a_1 > 0\},$$

où \mathcal{F}_0 désigne l'ensemble des fonctions continues périodiques telles que :

$$\int_0^{2\pi} f(t) \frac{dt}{(2\pi)} = 0.$$

Dans le chapitre 2, nous avons vu que ces contraintes d'identifiabilité apparaissent naturellement dans l'étude de la vraisemblance du modèle, et qu'au finale elles permettent de définir des estimateurs asymptotiquement efficace.

3.1.2 Estimation des paramètre pour le modèle (3.2).

La méthode d'estimation est celle décrite dans le chapitre 2. Elle est basée sur l'approximation de f^* par un polynôme trigonométrique de degré m_n . Afin que l'estimation soit asymptotiquement efficace, la suite m_n est croissante et tend vers l'infini

avec n . Les statistiques de test sont définies à partir d'un processus de contraste pour l'estimation du paramètre α^* . Or les critères développés au chapitre 2 ne sont pas les mieux indiqués pour définir la statistique de test (voir la section 3.2). La sous-section suivante présente de nouveaux processus de contraste, similaires à ceux étudiés dans le chapitre précédent, et qui sont effectivement utilisés pour définir les statistiques de test dans la section 3.2.

Estimation de α^* .

Dans le chapitre 2, on considère le critère (2.26) défini à partir des moindres carrés profilé :

$$L_n(\alpha) = \frac{1}{nJ} \sum_{i=1}^n \sum_{j=1}^J (Y_{i,j} - v_j)^2 - \sum_{1 \leq |l| \leq m_n} |\hat{c}_l(\alpha)|^2,$$

$$\text{où } \hat{c}_l(\alpha) = \frac{1}{nJ} \sum_{s=1}^n \sum_{j=1}^J a_j (Y_{sj} - v_j) e^{-il(t_s - \theta_j)} \quad \text{for } 1 \leq |l| \leq m_n, \quad (3.4)$$

et $(m_n)_n$ est une suite d'entiers tel que

$$2m_n < n, \quad \lim_{n \rightarrow \infty} m_n = +\infty, \quad \text{et } n \text{ est impair.} \quad (3.5)$$

Comme les points d'observation sont équidistants, la base de la transformée de Fourier discrète est orthogonale (2.9). Par conséquent les coefficients (\hat{c}_l) sont seulement fonction des paramètres $\theta = (\theta_j)_{j=1 \dots J}$ et $a = (a_j)_{j=1 \dots J}$. Par la suite, on s'autorise donc à noter $\hat{c}_l(\theta, a)$ à la place de $\hat{c}_l(\alpha)$.

A partir du critère L_n , on définit un nouveau processus de contraste M_n . Ce critère est légèrement similaire au critère (2.16) et donc aussi à celui proposé par Gamboa et al. (2005). Rappelons que la transformée discrète du modèle (2.11) transforme les données de la manière suivante :

$$d_{jl} = \frac{1}{n} \sum_{s=1}^n Y_{sj} e^{-ilt_s} = c_l^n(f_j^*) + \sigma^* \xi_{jl} / \sqrt{n}, \quad |l| \leq m_n \text{ et } j = 1 \dots J.$$

Le processus de contraste étudié ici est :

$$M_n(\alpha) = \frac{1}{J} \sum_{j=1}^J (d_{j0} - v_j)^2 + \sum_{1 \leq |l| \leq m_n} \sum_{j=1}^J |d_{jl} - a_j e^{-il\theta_j} \hat{c}_l(\theta, a)|^2. \quad (3.6)$$

On considère également le processus de contraste W_n qui diffère du précédent par l'introduction d'une suite de poids $(\delta_l)_{l \in \mathbb{Z}^*}$ de carré sommable :

$$W_n(\alpha) = \frac{1}{J} \sum_{j=1}^J (d_{j0} - v_j)^2 + \sum_{1 \leq |l| \leq m_n} \sum_{j=1}^J \delta_l^2 |d_{jl} - a_j e^{-il\theta_j} \hat{c}_l(\theta, a)|^2. \quad (3.7)$$

Comme dans la sous-section 2.2.2, il est nécessaire de supposer que l'une des hypothèses suivantes soient vérifiées,

(W1) les coefficients $c_1(f^*)$ et δ_1 sont non nuls.

ou

(W2) il existe $(l, p) \in \mathbb{Z}^2$ premiers entre eux tels que $c_l(f^*)$, $c_p(f^*)$, δ_l et δ_p soient non nuls.

Des théorèmes 2.3 et 2.1, on déduit les corollaires suivants sur la consistance des estimateurs définis à partir de M_n et W_n :

Corollaire 3.1. *Soit $\hat{\alpha}_n$ l'estimateur défini par :*

$$\hat{\alpha}_n = \arg \min_{\beta \in \mathcal{A}_3} M_n(\beta).$$

Si l'une des deux hypothèses (W1) ou (W2) est vérifiée, et que la fonction f^ et la suite m_n satisfont les hypothèses suivantes,*

$$\sum_{|l| > m} |c_l(f^*)| = o\left(\frac{1}{\sqrt{m}}\right), \quad (3.8)$$

$$\frac{m_n}{n} = o(1), \quad (3.9)$$

alors $\hat{\alpha}_n$ converge en probabilité vers α^ .*

La preuve est similaire au théorème 2.3. Remarquons que M_n converge uniformément en probabilité vers

$$M(\alpha) = \int_0^{2\pi} \frac{1}{J} \sum_{j=1}^J (f_j^*(t) - v_j)^2 \frac{dt}{2\pi} - \int_0^{2\pi} \left(\frac{1}{J} \sum_{j=1}^J a_j a_j^* f^*(t - \theta_j^* + \theta_j) \right)^2 \frac{dt}{2\pi},$$

tandis que L_n converge uniformément en probabilité vers $M(\alpha) + \sigma^{*2}$.

Par ailleurs, l'estimateur de v_j^* ($j = 1, \dots, J$) défini par les processus M_n et W_n est égal à d_{j0} . Il ne reste donc plus que $2J - 2$ paramètre à partir de la procédure de minimisation pour les critères M_n et W_n , à la place de $3J - 2$ pour le contraste L_n .

Corollaire 3.2. *Soit $\hat{\alpha}_n^W$ l'estimateur défini par :*

$$\hat{\alpha}_n^W = \arg \min_{\beta \in \mathcal{A}_3} W_n(\beta).$$

Si l'une des deux hypothèses (W1) ou (W2) est vérifiée, et que la fonction f^ et la suite de poids satisfont les hypothèses suivantes,*

$$\sum_{l \in \mathbb{Z}} |c_l(f^*)| < \infty, \quad \text{et} \quad \sum_{l \in \mathbb{Z}} \delta_l^2 l^2 < \infty, \quad (3.10)$$

alors $\hat{\alpha}_n^W$ converge en probabilité vers α^ .*

Soit ψ la \mathbb{L}^2 -fonction suivante :

$$\forall t \in \mathbb{R}, \quad \psi(t) = \sum_{l \in \mathbb{Z}} \delta_l e^{ilt}.$$

Le critère W_n converge uniformément en probabilité vers :

$$\int_0^{2\pi} \frac{1}{J} \sum_{j=1}^J (\psi * f_j^*(t) - v_j)^2 \frac{dt}{2\pi} - \int_0^{2\pi} \left(\frac{1}{J} \sum_{j=1}^J a_j a_j^* \psi * f^*(t - \theta_j^* + \theta_j) \right)^2 \frac{dt}{2\pi}.$$

Dans le corollaire 3.2, aucune hypothèse n'est posée sur le nombre de coefficients de Fourier m_n . En effet comme pour le processus de contraste (2.16), le corollaire est valide pour tout $m_n < (n/2)$, et donc pour $m_n = (n-1)/2$ (ce dernier choix n'est bien sûr pas possible avec le processus de contraste M_n .) Remarquons que l'hypothèse (3.8) signifie que la fonction f^* doit être au moins $1/2$ -Lipschitz pour assurer la consistance de $\hat{\alpha}_n$. Pour le corollaire 3.2, il est seulement exigé que la fonction f^* soit continue (hypothèse (3.10)). La séquence de poids régularise donc le critère et même temps que de contrôler le bruit.

Les propriétés asymptotiques des estimateurs $\hat{\alpha}_n$ et α_n^W découlent des théorèmes 2.2 et 2.4. Comme le gradient et la matrice hessienne de M_n sont identiques (à un petit $o_{\mathbb{P}}(1)$ près) au gradient et à la matrice hessienne de L_n , $\hat{\alpha}_n$ est donc asymptotiquement efficace.

Corollaire 3.3. *Supposons que l'une des deux hypothèses (W1) ou (W2) soit vérifiée et que :*

$$\sum_{l \in \mathbb{Z}} |l| |c_l(f^*)| < \infty, \quad (3.11)$$

$$m_n^4/n = o(1), \quad (3.12)$$

alors $\hat{\alpha}_n$ est asymptotiquement efficace et $\sqrt{n}(\hat{\alpha}_n - \alpha^*)$ converge en loi vers un vecteur Gaussien or $\mathcal{N}_{3J-2}(0, \sigma^{*2} H^{-1})$. H est la matrice suivante,

$$H = \begin{pmatrix} \|\partial f^*\|_{\mathbb{L}^2}^2 \left(D_a^2 - \frac{1}{J} V_a^2 V_a^{2T} \right) & 0 & 0 \\ 0 & \|f^*\|_{\mathbb{L}^2}^2 \left(I + \frac{1}{a_1^{*2}} a \ a^T \right) & 0 \\ 0 & 0 & I_J \end{pmatrix}.$$

où D_a est la matrice diagonale $\text{diag}(a_2^* \dots a_J^*)$, $a = (a_2^*, \dots, a_J^*)$ et $V_a = (a_2^{*2}, \dots, a_J^{*2})^T$ sont des vecteurs de \mathbb{R}^{J-1} .

Par contre, l'estimateur $\hat{\alpha}_n^W$ est asymptotiquement normal, mais pas efficace.

Corollaire 3.4. *Supposons que l'une des deux hypothèses (W1) ou (W2) soit vérifiée et que :*

$$\sum_{l \in \mathbb{Z}} |c_l(f^*)| < \infty, \quad \text{et} \quad \sum_{l \in \mathbb{Z}} \delta_l^2 l^2 < \infty,$$

alors $\sqrt{n}(\alpha_n^W - \alpha^*)$ converge en loi vers un vecteur gaussien $\mathcal{N}_{3J-2}(0, \sigma^{*2} \Gamma_W)$, où Γ_W est la matrice définie par

$$\Gamma_W = \begin{pmatrix} \frac{\sum_{l \in \mathbb{Z}} \delta_l^4 l^2 |c_l(f^*)|^2}{\sum_{l \in \mathbb{Z}} \delta_l^2 l^2 |c_l(f^*)|^2} (D_a^2 - \frac{1}{J} V_a V_a^T)^{-1} & 0 & 0 \\ 0 & \frac{\sum_{l \in \mathbb{Z}} \delta_l^4 |c_l(f^*)|^2}{\sum_{l \in \mathbb{Z}} \delta_l^2 |c_l(f^*)|^2} \left(I + \frac{1}{a_1^{*2}} a a^T \right)^{-1} & 0 \\ 0 & 0 & I_J \end{pmatrix}.$$

Comme précédemment, le corollaire 3.1 nécessite une régularité plus forte que le corollaire 3.4 : la fonction f^* est dérivable et sa dérivée est continue. Concernant le corollaire 3.4 toute la régularité a été reportée sur la suite de poids : on suppose seulement la continuité de f^* .

Estimation de f^*

Le corollaire 2.1 établit la convergence uniforme en probabilité de l'estimateur de la fonction f^* :

$$\hat{f}_n(t) = \sum_{1 \leq |l| \leq m_n} \hat{c}_l(\hat{\beta}_n) e^{ilt}, \quad \forall t \in \mathbb{R},$$

où $\hat{\beta}_n$ est un estimateur de α^* , en particulier $\hat{\beta}_n \in \{\hat{\alpha}_n, \hat{\alpha}_n^W\}$. Par définition de \hat{c}_l , cet estimateur est en fait la moyenne pondérée et rephasée des fonctions f_j^* .

Corollaire 3.5. *Supposons que $\hat{\beta}_n$ soit asymptotiquement normal et que*

$$\sum_{l \in \mathbb{Z}} |c_l(f^*)| < \infty, \quad \text{et} \quad \lim_{n \rightarrow \infty} \frac{m_n}{\sqrt{n}} = 0,$$

alors, \hat{f}_n converge en probabilité vers f^* pour la norme uniforme,

$$\sup_{t \in \mathbb{R}} \left| \hat{f}_n(t) - f^*(t) \right| = o_{\mathbb{P}}(1).$$

De plus, on a que $MISE_{f^*}(\hat{f}_n)$ converge vers zéro.

3.1.3 Estimation des paramètres pour le modèle général (3.1)

Afin de mettre en oeuvre la procédure bootstrap, il est nécessaire de définir des estimateurs consistants pour les fonctions de régression f_j^* et de la variance commune σ^{*2} .

Plus explicitement, on souhaite estimer les paramètres du modèle semi-paramétrique suivant :

$$Y_{i,j} = f_j^*(t_i) + \sigma^* \epsilon_{i,j}, \quad i=1 \dots n, \quad j=1 \dots J.$$

Le paramètre de nuisance est ici la fonction vectorielle (f_1^*, \dots, f_J^*) , et le paramètre fini-dimensionnel est σ^* . On estime ces paramètres en utilisant la vraisemblance profilée comme dans la sous-section 2.2.3.

La log-vraisemblance est égale à

$$l_n(\sigma, f_1, \dots, f_J) = \frac{-1}{2\sigma^2} \sum_{i=1}^n \sum_{j=1}^J (Y_{i,j} - f_j(t_i))^2 - \frac{nJ}{2} \log \sigma^2.$$

La première étape consiste à maximiser la log-vraisemblance sur le paramètre de nuisance. Pour cela, on approche les fonctions de régression par les estimateurs de leurs séries de Fourier tronquées. Les estimateurs des fonctions (f_j^*) sont alors :

$$\forall j=1 \dots J \quad \text{et} \quad \forall t \in \mathbb{R}, \quad \hat{f}_{j,n}(t) = \sum_{|l| \leq m_n} d_{j,l} e^{ilt},$$

où $(d_{j,l})$ est la transformée de Fourier discrète des données $(Y_{i,j})$, c'est-à-dire,

$$d_{j,l} = c_l^n(f_j^*) + \xi_{j,l}/\sqrt{n} \quad \text{pour} \quad |l| \leq m_n, \quad j = 1, \dots, J. \quad (3.13)$$

Le corollaire suivant est une conséquence du corollaire 2.1.

Corollaire 3.6. *Soit l'estimateur de σ^{*2} défini comme :*

$$\hat{\sigma}_n^2 = \frac{1}{nJ} \sum_{j=1}^J \sum_{i=1}^n \left(Y_{i,j} - \hat{f}_j(t_i) \right)^2. \quad (3.14)$$

Supposons que m_n/\sqrt{n} converge vers zéro et que,

$$\sum_{|l| > m} |c_l(f_j^*)| = o\left(\frac{1}{\sqrt{m}}\right) \quad \text{pour tout} \quad j = 1, \dots, J.$$

*Alors $\hat{\sigma}_n^2$ converge en probabilité vers σ^{*2} , et $(\hat{f}_{1,n} \dots \hat{f}_{J,n})$ converge en probabilité pour la norme uniforme vers (f_1^*, \dots, f_J^*) .*

Remarque 3.2. *Dans la pratique, le nombre de coefficients de Fourier m_n doit être relativement petit par rapport à n . Il faut un équilibre entre l'estimation de la variance et l'estimation des fonctions de régression.*

3.1.4 Application de la méthode d'estimation sur les ondes cérébrales P300

Les ondes P300 mesurent la capacité d'attention du patient. Ces ondes sont mesurées à partir de capteurs, en général placés sur la tête, qui enregistrent les ondes cérébrales et les transmettent, après décodage, à un ordinateur. Le procédé s'appuie le plus souvent sur l'électroencéphalographie (EEG) : toutes les activités humaines génèrent une activité cérébrale qui se manifeste par l'activation de zones distinctes dans le cerveau. En enregistrant ces faibles signaux, et en les associant aux actions concernées, il est possible de transformer une activité cérébrale en signal numérique et de le communiquer à un ordinateur.

Les ondes P300 résultent d'une bizarrerie du fonctionnement du cerveau : lorsqu'il est confronté à un événement surprenant ou inattendu, on constate une augmentation de la tension électrique du cerveau, 300 millisecondes après cet événement. On peut les simuler artificiellement de la manière suivante. On présente à l'utilisateur plusieurs choix, et quand celui qui lui convient apparaît, le simple fait de vouloir le sélectionner augmente la tension électrique du cerveau d'une manière plus aisément mesurable que son activité habituelle ne le permet.

Afin de déterminer l'efficacité d'un traitement, on observe sur chaque patient l'effet placebo (sans traitement) et l'effet du traitement. On a alors pour chaque patient deux ondes P300. Les scientifiques s'attendent en cas d'efficacité à observer une translation local au niveau du pic P300 sous l'effet du traitement.

Nous étudions ces deux effets sur onze patients. Les figures 3.1(a) et 3.1(d) représentent les onze signaux respectivement sous l'effet placebo et sous l'effet du traitement pour le capteur n°6. On implémente la procédure d'estimation présentée dans la section (3.1.2).

D'abord, les signaux placebo et les signaux du traitement sont étudiés simultanément dans le modèle suivant :

$$Y_{i,j} = a_j^* f^*(t_i - \theta_j^*) + \sigma^* \epsilon_{i,j}, \quad j = 1, \dots, 22, \quad i = 1, \dots, 127.$$

Les figures 3.1(b) et 3.1(e) représentent les estimateurs des fonctions de régression $f_j^*(\cdot) = a_j^* f^*(\cdot - \theta_j^*)$, $j = 1, \dots, J$ associés à chacun des deux effets. On remarque que la méthode d'estimation détecte une forme commune. Elle est plus aplatie pour les ondes associées au traitement.

Mais l'étude suivante jette un doute sur la validité du modèle. En effet, si on considère séparément les signaux placebo et les signaux du traitement, on a alors deux études avec $J = 11$ signaux. Pour chacune de ces deux études, on considère le

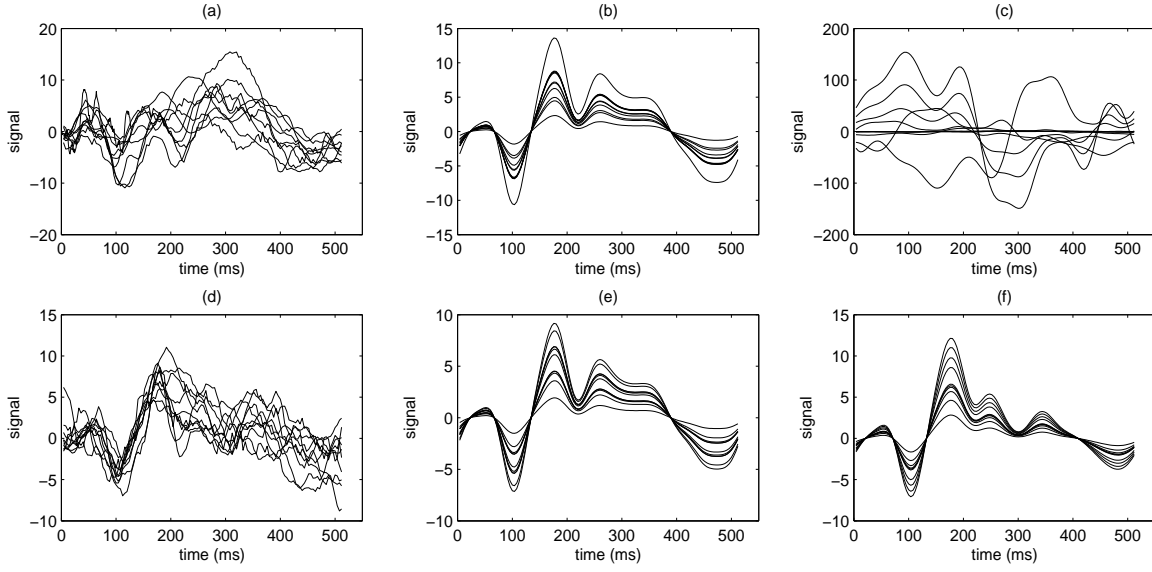


FIG. 3.1: Les ondes P300 associées respectivement à l'effet placebo (a) et à l'effet du traitement (d). L'estimation des fonctions de régression dans le modèle avec $J = 22$ unités pour les ondes placebo (b) et les ondes traitements (e). L'estimation des fonctions de régression dans le modèle avec $J = 11$ courbes en considérant uniquement les ondes placebo (c) et les ondes traitement (f).

modèle suivant,

$$Y_{i,j} = a_j^* f^*(t_i - \theta_j^*) + \sigma^* \epsilon_{i,j}, \quad j = 1, \dots, 11 \quad i = 1, \dots, 127.$$

La figure 3.1(c) représente les estimateurs des fonctions de régression pour le modèle avec les onze signaux placebo. La figure 3.1(f) représente les estimateurs des fonctions de régression pour le modèle avec les onze signaux de traitement. Alors qu'une forme commune est toujours détectée avec le modèle considérant uniquement les signaux de traitement, aucune forme commune n'est visiblement détectée pour le modèle considérant uniquement les signaux placebo. Les prochaines sections cherchent à fournir des procédures de test afin de tester la validité du modèle.

3.2 Test d'adéquation du modèle

Dans cette section, le but est de tester à partir du modèle (3.1) la validité de l'hypothèse nulle H_0 ,

il existe $\alpha \in \mathcal{A}_3$ et une 2π -périodique fonction f telle que pour tout $j = 1 \dots J$,

$$f_j^*(\cdot) = a_j f(\cdot - \theta_j) + v_j,$$

contre l'hypothèse alternative H_1 ,

il n'existe aucun $\alpha \in \mathcal{A}_3$ et aucune fonction f de telle sorte que les fonctions de régression f_j^* soient liées entre elles par l'équation (3.2).

Une première idée serait de prendre le processus de contraste L_n défini au chapitre 2 (et dont la définition a été rappelée dans la section précédente) et de considérer la statistique de test suivante,

$$T_n^L = nJ \inf_{\alpha \in \mathcal{A}_3} L_n(\alpha).$$

Sous l'hypothèse nulle cette statistique diverge. Nous présentons ici une heuristique. Un développement de Taylor nous permet de récrire la statistique de test sous la forme suivante,

$$T_n^L = J \left(\sqrt{n} \nabla L_n(\alpha^*) \right)^T H^{-1} \sqrt{n} \nabla L_n(\alpha^*) + Jn L_n(\alpha^*) + o_{\mathbb{P}}(1).$$

La proposition (2.5) montre que $\sqrt{n} \nabla L_n(\alpha^*)$ converge en loi vers un vecteur gaussien. Ainsi le premier terme est uniformément tendue. D'autre part, L_n converge uniformément en probabilité vers $M + \sigma^{*2}$ (lemme 2.6). Comme $M(\alpha^*)$ est nulle et $\sigma^* \neq 0$, le second terme tend vers l'infini.

Les processus de contraste M_n et W_n sont définis de telle sorte que leurs statistiques de test associées convergent en loi sous l'hypothèse nulle.

Cette section est divisée en trois sous-sections. D'abord, les statistiques des test sont définies et étudiées sous l'hypothèse nulle et l'hypothèse alternative. Ensuite, on décrit la procédure bootstrap afin de mettre en oeuvre le test statistique. Puis, on présente quelques simulations.

3.2.1 Les statistiques de test.

Pour l'instant, on considère le modèle où les écarts-type σ_j^* , $j = 1 \dots J$, sont tous égaux. Le modèle, où les écarts-type ne sont pas tous égaux, est discuté dans le remarque 3.3. A partir des processus de contraste M_n et W_n , on construit deux statistiques de test notées respectivement T_n et T_n^W . Le théorème 3.1 étudie la statistique de test construite à partir du processus M_n .

Théorème 3.1. *Soit T_n la statistique suivante,*

$$T_n = Jn \inf_{\alpha \in \mathcal{A}_3} M_n(\alpha) / \hat{\sigma}_n^2.$$

Supposons que les hypothèses du corollaire 3.3 soient vérifiées et que les fonctions f_j^ satisfont l'hypothèse suivante :*

$$\sum_{l \in \mathbb{Z}} |l| |c_l(f_j^*)| < +\infty, \quad j = 1, \dots, J.$$

Alors,

- sous H_0 , T_n/m_n converge en probabilité vers $2J - 1$;
- sous H_1 , T_n/m_n tend vers l'infini.

Le théorème 3.1 n'est pas applicable dans la pratique. En effet, le résultat est vrai si le nombre de coefficients de Fourier m_n est grand. Par conséquent, l'hypothèse (3.12) implique que la taille des données n soit excessivement grande. C'est pourquoi on considère le critère W_n qui est moins restrictif sur le nombre de coefficients de Fourier m_n .

Corollaire 3.7. Soit T_n^W la statistique définie comme

$$T_n^W = Jn \inf_{\alpha \in A_3} W_n(\alpha) / \hat{\sigma}_n^2.$$

Supposons que la suite $(\delta_l)_l$ est positive, et que les hypothèses du corollaire 3.4 soient vérifiées, ainsi que

$$\frac{m_n}{\sqrt{n}} = o(1), \quad \text{et} \quad \sum_{|l| > m} |c_l(f_j^*)| = o\left(\frac{1}{\sqrt{m}}\right), \quad j = 1, \dots, J.$$

Alors,

- sous H_0 , T_n^W converge en loi;
- sous H_1 , T_n^W tend vers l'infini.

D'après la preuve du théorème 2.5 de Bayomog et al. (1996), on s'attend à ce que la loi limite de T_n^W sous H_0 soit une somme pondérée de variables indépendantes suivant une loi de chi-deux. Ici la loi limite est la somme de deux variables qui ne sont pas indépendantes. La première est une somme pondérée de variables indépendantes suivant une loi du chi-deux, cette variable dépend seulement de la suite de poids. La seconde est une forme quadratique d'un vecteur gaussien, elle dépend de la forme commune et du paramètre d'amplitude a^* . En fait, la loi limite est trop compliquée pour espérer établir un procédure plug-in (voir la remarque 3.4).

Ainsi, on utilise la statistique T_n à la place de T_n/m_n , et on utilise les résultats asymptotiques du corollaire 3.7. À cause de la complexité de la loi limite de T_n^W sous l'hypothèse nulle, la méthode bootstrap décrite dans la sous-section suivante permet d'approcher les quantiles de cette loi et de mettre en oeuvre le test statistique.

Remarque 3.3. Considérons le modèle où les écarts-type (σ_j^*) , $j = 1 \dots J$, ne sont pas tous égaux. Dans ce modèle, les processus de contrastes M_n et W_n permettent également la définition d'estimateurs consistants du paramètre α^* . Aussi, on peut considérer les

statistiques de test suivantes,

$$nJ \inf_{\alpha \in \mathcal{A}_3} M_n(\alpha) \quad \text{et} \quad nJ \inf_{\alpha \in \mathcal{A}_3} W_n(\alpha).$$

Ces statistiques de test se comportent asymptotiquement de la même manière que les statistiques T_n et T_n^W . Néanmoins, la distribution limite sont fonctions des écart types (σ_j^*) , $j = 1 \dots J$. De plus, une simple normalisation effectuée pour les statistiques T_n et T_n^W n'est pas possible.

3.2.2 La méthode bootstrap

Le bootstrap est une procédure pour estimer (approcher) une fonctionnelle d'une distribution d'une statistique. Efron (1979) a montré que dans des modèles complexes où l'estimateur plug-in est trop compliqué pour être calculé, on peut approcher cette fonctionnelle par rééchantillonnage (méthodes de Monte-Carlo).

Soit \mathbb{F} la distribution limite de T_n^W sous l'hypothèse nulle. On souhaite estimer le quantile de la loi \mathbb{F} afin de mettre en oeuvre le test. Beran (1988) et Fisher et Hall (1990) ont suggéré que pour qu'un test bootstrap fonctionne bien, la statistique de test doit être un pivot asymptotique. C'est-à-dire que la distribution asymptotique \mathbb{F} ne doit pas dépendre d'un des paramètres. Cependant Horowitz (1995) a observé que le bootstrap peut aussi être appliqué à des statistiques qui ne sont pas asymptotiquement des pivots (par exemple pour des coefficients de régression) cependant il n'apporte pas une approximation d'un ordre supérieur à celle de la théorie asymptotique. Divisé le processus de contraste W_n par l'estimateur de σ^{*2} permet que \mathbb{F} ne dépend pas de la variance σ^{*2} . Mais elle dépend de la forme commune de façon assez compliquée (voir la remarque 3.4), une statistique de pivot est alors difficilement constructible. Néanmoins, les simulations illustrent les bons résultats de la méthode (voir la sous-section 3.2.3).

L'idée du bootstrap est de générer un grand nombre d'échantillons bootstrap qui suit l'hypothèse nulle. Pour le modèle de régression considéré, on se réfère à la méthode développée par Efron et Tibshirani (1993). Les auteurs ont développé une méthode afin de générer des échantillons bootstrap pour un modèle de régression linéaire. Or le fait d'approcher les fonctions de régression par leurs séries de Fourier nous ramène en quelque sorte à ce modèle. On calcule donc les résidus sous l'hypothèse du modèle général (3.1),

$$\hat{\epsilon}_{i,j} = Y_{i,j} - \hat{f}_{j,n}(t_i) \quad \text{for } j = 1 \dots J, \quad i = 1 \dots n,$$

où $\hat{f}_{j,n}$ sont les estimateurs des fonctions de régression pour le modèle général (voir le corollaire 3.6). Afin que les résidus soient bien un échantillon de variables centrées,

on effectue le centrage suivant,

$$\tilde{\epsilon}_{i,j} = \hat{\epsilon}_{i,j} - \sum_{i_0=1}^n \sum_{j_0=1}^J \hat{\epsilon}_{i_0 j_0} / (nJ) \quad \text{for } j = 1 \dots J, i = 1 \dots n,$$

On génère alors un échantillon bootstrap $\{(\epsilon_{i,j}^*), i = 1 \dots n, j = 1 \dots J\}$ des résidus centrés (voir chapitre 1, section 1.3). Pour cela, on procède à nJ tirages avec remise des résidus centrés $\tilde{\epsilon}_{i,j}$. L'échantillon bootstrap des données sous l'hypothèse nulle est obtenu de la manière qui suit :

$$\hat{Y}_{i,j}^* = \hat{a}_j \hat{f}_n(t_i - \hat{\theta}_{j,n}) + \hat{v}_{j,n} + \tilde{\epsilon}_{i,j}^* \quad \text{for } j = 1 \dots J, i = 1 \dots n, \quad (3.15)$$

où $((\hat{\theta}_{j,n})_{j=1 \dots J}, (\hat{a}_{j,n})_{j=1 \dots J}, (\hat{v}_{j,n})_{j=1 \dots J})$ est l'estimateur de α^* défini dans le corollaire 3.1 ou 3.2, et \hat{f}_n est l'estimateur de la forme commune f^* défini dans le corollaire 3.5. Pour cet échantillon bootstrap des données sous l'hypothèse nulle, on calcule la statistique de test,

$$\hat{T}^* = nJ \inf_{(\theta, a, v) \in \mathcal{A}_3} \hat{W}_n^*(\theta, a, v) / \hat{s}^{2*},$$

où \hat{W}_n^* est le critère défini à partir de l'échantillon bootstrap (3.15), \hat{s}^{2*} est l'estimateur de la variance (voir corollaire 3.6) calculé à partir de l'échantillon bootstrap (3.6).

La procédure est répétée B fois. Soit \hat{T}_b^* la statistique de test calculée à partir de la $b^{\text{ème}}$ échantillon bootstrap pour $b = 1 \dots B$. Ainsi $(\hat{T}_b^*)_{b=1 \dots B}$ est un échantillon bootstrap de la distribution \mathbb{F} . Il suffit alors de considérer la statistique d'ordre pB de cet échantillon, que nous notons $T_{(pB)}^*$, pour obtenir une approximation de la quantile d'ordre $p \in [0, 1]$ de \mathbb{F} . On rejette alors l'hypothèse nulle au niveau $1 - p$ si,

$$T_n > T_{(pB)}^*.$$

3.2.3 Les simulations

Pour chaque modèle, le tableau 3.1 affiche les coefficients de Fourier normalisé ainsi que le processus de contraste associés.

Nous avons vérifié la validité de notre test à un niveau de confiance de 95% pour les trois modèles décrits dans la table 3.1. Les résultats sont résumés dans les figures 3.2, 3.3 et 3.4.

Concernant la figure 3.2, les données sont générées à partir du modèle suivant :

$$\begin{cases} Y_{i,1} = a_1^* f^*(t_i - \theta_1^*) + v_1^* + \sigma^* \epsilon_{i1} & i = 1 \dots n, \\ Y_{i,2} = a_2^* f^*(\beta t_i - \theta_2^*) + v_2^* + \sigma^* \epsilon_{i2} & i = 1 \dots n, \end{cases}$$

L'hypothèse H_0 .	Les coefficients de Fourier $\hat{c}_b l \leq m_n$	Le processus M_n .
$a_j^* f^*(\cdot - \theta_j^*) + \nu_j^*$	voir (3.4)	voir (3.6)
$a_j^* f^*(\cdot - \theta_j^*)$	$\hat{c}_l(\theta, a) = \sum_{j=1}^J \frac{a_j e^{il\theta_j} d_{jl}}{J}$	$\sum_{ l \leq m_n} \sum_{j=1}^J \left d_{jl} - a_j e^{-il\theta_j} \hat{c}_l(\theta, a) \right ^2$
$f^*(\cdot - \theta_j^*)$	$\hat{c}_l(\theta, a) = \sum_{j=1}^J \frac{e^{il\theta_j} d_{jl}}{J}$	$\sum_{ l \leq m_n} \sum_{j=1}^J \left d_{jl} - e^{-il\theta_j} \hat{c}_l(\theta, a) \right ^2$

TAB. 3.1: Correspondance entre l'hypothèse nulle à tester et le processus de contraste associé.

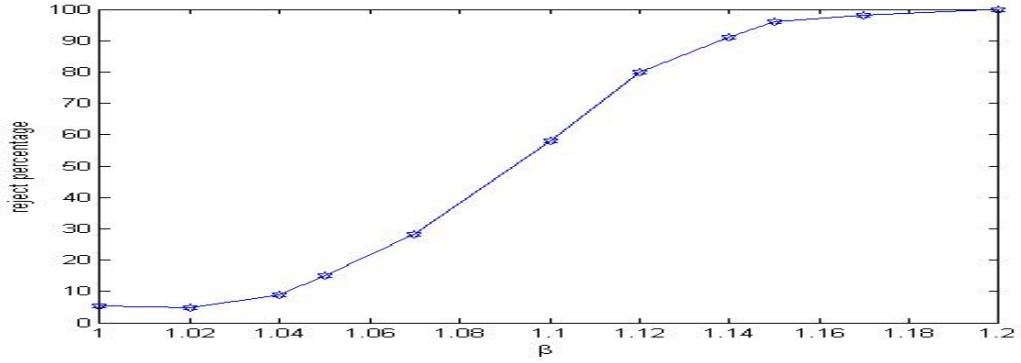


FIG. 3.2: Puissance du test pour l'hypothèse nulle " $f_j^* = a_j^* f^*(\cdot - \theta_j^*) + \nu_j^*$ " avec $J = 2$ en fonction de la variable β . Chaque point est calculé à partir de 1000 procédures de test, et les paramètres sont $(\theta_1^*, \theta_2^*) = (0, 0.8)$, $(a_1^*, a_2^*) = (0.75, 1.19)$, $(\nu_1^*, \nu_2^*) = (-1, 0.5)$, $\sigma^* = 0.8$, et $n = 201$.

où β appartient à l'intervalle $[1, 1.2]$, les autres paramètres étant fixés (voir la figure 3.2). En particulier, pour toutes les simulations, la forme commune est identique et elle est égale à :

$$f^*(t) = 20 \left(1 - t/(2\pi)\right) t/(2\pi) - 10/3, \quad t \in \mathbb{R}.$$

La figure 3.2 représente l'évolution du pourcentage de rejet de l'hypothèse nulle (3.2) en fonction de la variable β . Quand H_0 est vraie (c'est-à-dire $\beta = 1$), le pourcentage de rejet est proche de la valeur espérée 5%. De plus, il est visible que la procédure de test rejette largement l'hypothèse nulle même si la modification de la forme de f_2^* est petite en comparaison du niveau de bruit. Par exemple, le pourcentage de rejet est 57.9% quand $\beta = 1.1$.

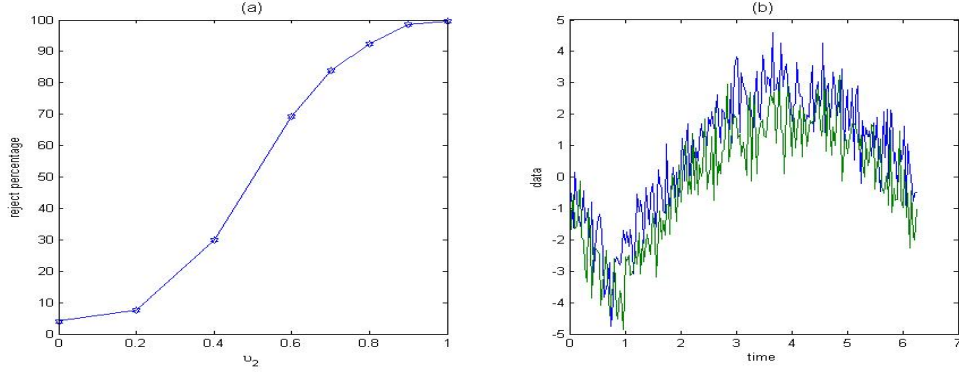


FIG. 3.3: (a) Puissance du test pour l'hypothèse nulle " $f_j^* = a_j^* f^*(\cdot - \theta_j^*)$ " avec $J = 3$ en fonction de v_2 . Chaque point est calculé à partir de 1000 procédures de test et les paramètres sont $(\theta_1^*, \theta_2^*, \theta_3^*) = (0, 0.27, 0.83)$, $(a_1^*, a_2^*, a_3^*) = (0.75, 1.19, 1)$, $\sigma^* = 0.8$, et $n = 201$. (b) Les données sont générées à partir de $Y_{i,1} = a_2^* f^*(t_i - \theta_2^*) + \sigma^* \epsilon_{i1}$, et $Y_{i,2} = a_2^* f^*(t_i - \theta_2^*) + 0.8 + \sigma^* \epsilon_{i2}$, $i = 1, \dots, n$.

Concernant la figure 3.3, le modèle considéré est le suivant :

$$\begin{cases} Y_{i,j} = a_j^* f^*(t_i - \theta_j^*) + \sigma^* \epsilon_{i,j} & i = 1 \dots n, j = 1, 3 \\ Y_{i,2} = a_2^* f^*(t_i - \theta_2^*) + v_2 + \sigma^* \epsilon_{i2} & i = 1 \dots n, \end{cases}$$

où v_2 appartient à l'intervalle $[0, 1]$, les autres paramètres sont fixés (voir la figure 3.3) et la fonction f^* est définie précédemment. La figure 3.3(a) représente l'évolution du pourcentage de rejet de l'hypothèse nulle $f_j^* = a_j^* f^*(\cdot - \theta_j^*)$ en fonction de la variable v_3 . Quand H_0 est vraie (c'est-à-dire quand $v_2 = 0$), le pourcentage de rejet est proche de la valeur attendue (5%). De plus, le pourcentage de rejet est très élevé (92.3%) pour $v_2 = 0.8$ alors que les données $(Y_{i2})_{i=1 \dots n}$ sont très proche des données générées sous l'hypothèse nulle (voir la figure 3.3(b) pour une comparaison graphique de ces deux ensemble de données),

$$Y_{i2} = a_2^* f^*(t_i - \theta_2^*) + \sigma^* \epsilon_{i2} \quad i = 1 \dots n.$$

Concernant la figure 3.4, les données sont générées à partir du modèle suivant :

$$\begin{cases} Y_{i,j} = f^*(t_i - \theta_j^*) + \sigma^* \epsilon_{i,j} & i = 1 \dots n, j = 1, 3 \\ Y_{i,2} = a_2 f^*(t_i - \theta_2^*) + \sigma^* \epsilon_{i2} & i = 1 \dots n, \end{cases}$$

où a_2 appartient à l'intervalle $[1, 1.4]$, et les autres paramètres sont fixés. La figure 3.4(a) représente l'évolution du pourcentage de rejet de l'hypothèse nulle $f_j^* = f^*(\cdot - \theta_j^*)$ en fonction de la variable a_2 . Quand H_0 est vraie (c'est-à-dire $a_2 = 0$), le

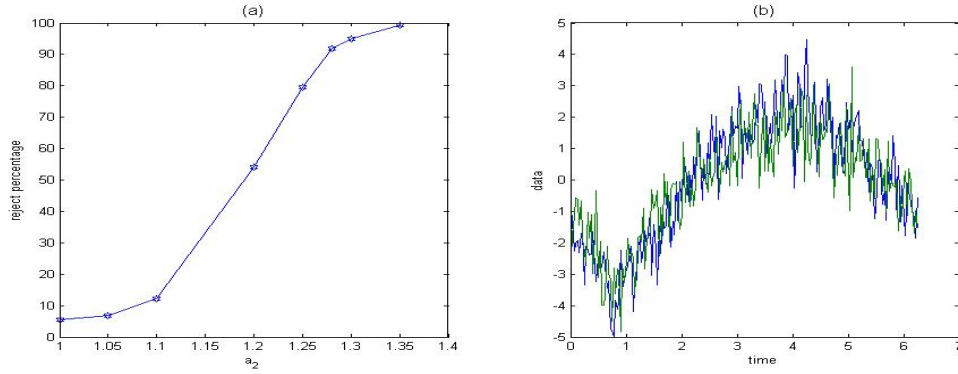


FIG. 3.4: (a) Puissance du test pour l'hypothèse nulle $f_j^* = f^*(\cdot - \theta_j^*)$ avec $J = 3$ en fonction de a_2 . Chaque point est calculé à partir de 1000 procédures de test et les paramètres sont $(\theta_1^*, \theta_2^*, \theta_3^*) = (0, 0.27, 0.83)$, $\sigma^* = 0.8$, et $n = 201$. (b) Les données sont générées à partir de $Y_{i,1} = f(t_i - \theta_2^*) + \sigma^* \epsilon_{i1}$, et $Y_{i,2} = 1.3f(t_i - \theta_2^*) + \sigma^* \epsilon_{i2}$ $i = 1, \dots, n$.

pourcentage de rejet est proche de 5% comme espéré. De plus le pourcentage de rejet est 95% pour $v_2 = 1.3$, alors que la différence entre l'hypothèse nulle est négligable en comparaison du niveau de bruit (voir la figure 3.4(b)).

Application aux ondes P300.

Considérons d'abord les ondes de traitement et les ondes placebo simultanément. On a au total $J = 22$ signaux. La statistique de test est égale à 2652, alors que l'approximation bootstrap du 95^{ème} percentile est 27.230. Les données n'adhèrent donc pas au modèle.

Comme on l'a vu dans la section 3.1.4, si on considère les onze ondes placebo seules, elles n'adhèrent pas au modèle. Par conséquent, on applique seulement la procédure de test sur les onze ondes associées au traitement. La statistique de test est égale à 784.4004, alors que l'approximation bootstrap du 95^{ème} percentile est 25.61. On rejette aussi l'hypothèse nulle. Néanmoins, la valeur de la statistique de test est ici plus de trois fois moins grande que celle où l'on considère les signaux placebo et de traitement simultanément. On conclut que les réponses des patients sous traitement sont plus homogènes que sans traitement.

Pour conclure, les exemples montrent que même si un nombre petit de signaux, parmi les J considérés, n'avaient pas une forme commune avec les autres signaux, alors l'estimation de la forme commune en était très peu affectée. Par contre, une légère différence dans la forme d'un des signaux est suffisante pour que la procédure de test rejette l'hypothèse nulle. La procédure de test est donc très sensible.

3.2.4 Preuves

Preuve du théorème 3.1. Le processus de contraste peut se récrire comme la somme des deux fonctions suivantes

$$M_n^1(v) = \sum_{j=1}^J |d_{j0} - v_j|^2 / J \quad \text{et} \quad M_n^2(\theta, a) = \sum_{1 \leq |l| \leq m_n} \sum_{j=1}^J |d_{jl} - a_j e^{-il\theta_j} \hat{c}_l(\theta, a)|^2 / J,$$

où $v = (v_1, \dots, v_J)$, $\theta = (0, \theta_2, \dots, \theta_J)$, et $a = (\{J - \sum_{j=2}^J a_j^2\}^{1/2}, a_2, \dots, a_J)$. Supposons que H_0 soit vraie. Alors le minimum est atteint au point $\hat{\alpha}_n = (\hat{\theta}_n, \hat{a}_n, \hat{v}_n)$ par définition,

$$\hat{\sigma}_n^2 T_n = n J M_n^2(\hat{\gamma}_n) = -n J (M_n^2(\gamma^*) - M_n^2(\hat{\gamma}_n)) + n J M_n^2(\gamma^*),$$

où γ^* , $\hat{\gamma}_n$ sont respectivement les paramètres (θ^*, a^*) et $(\hat{\theta}_n, \hat{a}_n)$. Par un développement de Taylor en $\hat{\gamma}_n$, il existe $\tilde{\gamma}_n$ appartenant à $\{\gamma : \|\gamma - \gamma^*\| < \|\hat{\gamma}_n - \gamma^*\|\}$ tel que :

$$2\hat{\sigma}_n^2 T_n = -J\sqrt{n} {}^t(\hat{\gamma}_n - \gamma^*) \nabla^2 M_n^2(\tilde{\gamma}_n) \sqrt{n}(\hat{\gamma}_n - \gamma^*) + 2Jn M_n^2(\gamma^*).$$

D'après la proposition (2.6), la matrice hessienne $\nabla^2 M_n$ converge uniformément en probabilité vers une matrice $-2H_\gamma/J$ sur $\{\gamma : \|\gamma - \gamma^*\| < \|\hat{\gamma}_n - \gamma^*\|\}$. Par le corollaire 3.3, on déduit alors que

$$\hat{\sigma}_n^2 T_n = Jn M_n^2(\gamma^*) + \mathcal{O}_{\mathbb{P}}(1).$$

Comme m_n tend vers l'infini avec n , on obtient que

$$\begin{aligned} T_n/m_n &= Jn M_n^2(\gamma^*)/m_n + o_{\mathbb{P}}(1) \\ &= \frac{\sigma^{*2}}{\hat{\sigma}_n^2} \left(\sum_{1 \leq |l| \leq m_n} \sum_{j=1}^J |\xi_{jl}|^2 / m_n - J \sum_{1 \leq |l| \leq m_n} |\xi_l(\gamma^*)|^2 / m_n \right) + o_{\mathbb{P}}(1), \end{aligned}$$

où $\xi_l(\gamma^*) = \sum_{j=1}^J a_j^{*2} \xi_{jl} e^{-il\theta_j^*} / J$. Puisque l'estimateur $\hat{\sigma}_n^2$ converge en probabilité vers σ^* (corollaire 3.6), la loi forte des grands nombres implique que T_n/m_n converge en probabilité vers $2J - 2$.

Supposons maintenant que H_1 soit vraie. Soit $\gamma = (\theta, a)$ tel que $(\gamma, \hat{v}_n) \in \mathcal{A}_3$. Après calcul on a,

$$M_n(\gamma, \hat{v}_n) = n \sum_{1 \leq |l| \leq m_n} \sum_{j=1}^J \left| c_l^n(f_j^*) - a_j e^{-il\theta_j} \sum_{k=1}^J a_k e^{il\theta_k} c_l^n(f_k^*) / J \right|^2 + o_{\mathbb{P}}(\sqrt{n}).$$

Si les fonctions de régression f_j^* satisfont les hypothèses du théorème, il est facile de montrer que :

$$n J M_n(\gamma, \hat{v}_n) = n J M^2(\gamma) + o(1) + \mathcal{O}_{\mathbb{P}}(\sqrt{n}/m_n),$$

$$\text{où } M^2(\gamma) = \sum_{l \in \mathbb{Z}^*} \sum_{j=1}^J \left| c_l(f_j^*) - a_j e^{-il\theta_j} \frac{1}{J} \sum_{k=1}^J a_k e^{il\theta_k} c_l(f_k^*) \right|^2 / J.$$

Donc si la borne inférieure de M^2 est non nulle, on déduit que JnT_n^*/m_n tend vers l'infini. L'espace d'optimisation étant compact, M^2 atteint sa borne en un point γ (M^2 est continue). Ainsi $M^2(\gamma)$ est égal à zéro si, et seulement si, il existe une suite sommable (λ_l) telle que pour tout $l \in \mathbb{Z}^*$ et $j = 1 \dots J$:

$$c_l(f_j^*) = a_j \lambda_l e^{-il\theta_j}.$$

Autrement dit, la borne inférieure de M^2 est nulle si, et seulement si H_0 est vraie. \square

Preuve du corollaire 3.7. Le processus de contraste W_n peut se récrire comme la somme des deux processus suivants :

$$W_n^1(v) = M_n^1(v) \quad \text{et} \quad W_n^2(\theta, a) = \sum_{1 \leq |l| \leq m_n} \sum_{j=1}^J \delta_l^2 |d_{jl} - a_j e^{-il\theta_j} \hat{c}_l(\theta, a)|^2 / J.$$

L'étude de la statistique de test sous l'hypothèse alternative est identique à l'étude faite dans la preuve du théorème 3.1.

Étudions la statistique de test sous H_0 . Le minimum de W_n est atteint en $\hat{\alpha}_n^W = (\hat{\theta}_n^w, \hat{a}_n^w, \hat{v}_n)$. La statistique de test est donc égale à :

$$\hat{\sigma}_n T_n^W = JnW_n^2(\hat{\gamma}_n^w) = -Jn(W_n^2(\gamma^*) - W_n^2(\hat{\gamma}_n^W)) + JnW_n^2(\gamma^*),$$

où γ^* , $\hat{\gamma}_n^W$ représente respectivement (θ^*, a^*) et $(\hat{\theta}_n^W, \hat{a}_n^W)$. Par un développement de Taylor en $\hat{\gamma}_n^W$, il existe $\tilde{\gamma}_n \in \{\gamma : \|\gamma - \gamma^*\| < \|\hat{\gamma}_n^W - \gamma^*\|\}$ tel que :

$$2\hat{\sigma}_n^2 T_n = -J\sqrt{n} (\hat{\gamma}_n^W - \gamma^*)^T \nabla^2 W_n^2(\tilde{\gamma}_n) \sqrt{n} (\hat{\gamma}_n^W - \gamma^*) + 2JnW_n^2(\gamma^*). \quad (3.16)$$

De façon identique à la proposition 2.6, la matrice hessienne $\nabla^2 W_n$ converge uniformément en probabilité vers H_γ^W sur $\{\gamma : \|\gamma - \gamma^*\| < \|\hat{\gamma}_n - \gamma^*\|\}$. D'autre part, les fluctuations de l'estimateur $\hat{\gamma}_n$ sont fonctions du gradient de W_n^2 (par un développement de Taylor),

$$\sqrt{n} (\hat{\gamma}_n^W - \gamma^*) = \sigma^* (H_\gamma^W)^{-1} \nabla W_n^2(\gamma^*) + o_{\mathbb{P}}(1)$$

où $\nabla W_n^2(\gamma^*)$ convergent en loi vers un vecteur gaussien centré de matrice de covariance :

$$\begin{pmatrix} \sum_{1 \leq |l| \leq m_n} \delta_l^4 |lc_l(f)|^2 (D^2 - \frac{1}{J} V_a V_a^T) & 0 \\ 0 & \sum_{1 \leq |l| \leq m_n} \delta_l^4 |c_l(f)|^2 \left(I + \frac{1}{a_1^{*2}} a a^T \right) \end{pmatrix}.$$

D'autre part, $JnW_n^W(\gamma^*)$ est une somme pondérée de chi-deux qui converge dans $\mathbb{L}^2(\mathbb{P})$:

$$JnW_n^2(\gamma^*) = \sigma^{*2} \sum_{1 \leq |l| \leq m_n} \delta_l^2 \left(\sum_{j=1}^J |\xi_{jl}|^2 - J |\xi_l(\gamma^*)|^2 \right).$$

Ainsi, en reprenant (3.16), on a

$$2\hat{\sigma}_n^2 T_n = -J\sqrt{n} \left(\nabla W_n^2(\gamma^*) \right)^T \nabla^2 W_n^2(\tilde{\gamma}_n) \sqrt{n} \left(\nabla W_n^2(\gamma^*) \right) + 2JnW_n^2(\gamma^*) + o_{\mathbb{P}}(1).$$

On conclut que la statistique de test converge en loi. \square

Remarque 3.4. À cause de la dépendance des variables $JnW_n^2(\gamma^*)$ et ∇W_n^2 , il est difficile de déterminer exactement la distribution limite sous l'hypothèse nulle de la statistique T_n^W . Par exemple pour $J = 2$, la fonction caractéristique de la loi jointe $(G_n^W, JnW_n^2(\gamma^*))$ converge vers

$$\phi(t_\theta, t_a, u) = \phi_Z(t_\theta, t_a) \phi_Y(u) \exp \left(\sum_{l \neq 0} \frac{-i\delta_l^6 u^2 |c_l(f)|^2}{1 - i\delta_l^2 u^2} \left(\frac{t_\theta^2}{2} l^2 + t_a^2 \right) a_2^{*2} a_1^{*2} (1 - a_2^{*2}) \right), \quad (3.17)$$

où Z et Y sont respectivement les lois limites de ∇W_n^2 et $JnW_n^2(\gamma^*)$.

3.3 Test sur l'homogénéité des données

Dans cette section, le but est de construire une procédure statistique afin de tester à partir du modèle général (3.1) si les fonctions de régression f_j^* sont homogènes dans le sens suivant :

$$\text{pour tout } l \in \mathbb{Z}, \text{ pour tout } (j, k) \in \{1 \dots J\}^2, |c_l(f_j^*)| = |c_l(f_k^*)|. \quad (3.18)$$

À l'origine de ce travail, on voulait construire un test d'adéquation pour l'hypothèse suivante,

$$\forall j = 2, \dots, J, \quad \exists \theta_j \in \mathbb{R} \quad \text{tel que} \quad f_j(\cdot) = f_1(\cdot - \theta_j). \quad (3.19)$$

Une première approche était de considérer la statistique de test définie à partir d'un processus de contraste dérivé de (3.7) (cf. le tableau 3.1). Or, dans la section précédente, nous avons montré que la loi limite sous l'hypothèse nulle était trop compliquée. En effet, cette loi dépend des coefficients de Fourier de la forme commune de telle façon qu'il est difficile d'envisager un estimateur plug-in des quantiles de cette loi. Cette difficulté a été surmonté par l'utilisation d'un algorithme bootstrap.

Ici nous envisageons une solution différente. Au lieu de considérer tous les coefficients de Fourier en même temps, on les étudie un par un en se demandant si pour chaque $l \in \mathbb{Z}$ il existe $(\theta_{j,l})_{j=2 \dots J} \in \mathbb{R}^{J-1}$ tel que :

$$c_l(f_j^*) = c_l(f_1^*) e^{il\theta_{j,l}} \quad \forall j = 2, \dots, J.$$

Pour chaque coefficient de Fourier, on construit donc une statistique de test. Ensuite, ces statistiques de test sont agrégées afin de tester l'hypothèse (3.18). L'avantage est que la loi limite est asymptotiquement un pivot : elle ne dépend plus de la forme commune. Par contre, on perd au niveau de la précision de l'hypothèse nulle à tester : on ne teste plus,

$$\exists \theta = (\theta_j)_{j=2 \dots J} \in \mathbb{R}^{J-1} \quad \text{tel que} \quad \forall l \in \mathbb{Z}, \forall j \in \{2 \dots J\}, \quad c_l(f_j^*) = c_l(f_1^*) e^{il\theta_j},$$

mais, on teste l'hypothèse nulle (3.18) qui peut se récrire ainsi :

$$\forall l \in \mathbb{Z}, \exists \theta = (\theta_{j,l})_{j=2 \dots J} \in \mathbb{R}^{J-1} \quad \text{tel que} \quad \forall j \in \{2 \dots J\}, \quad c_l(f_j^*) = c_l(f_1^*) e^{il\theta_j}.$$

L'hypothèse (3.18) est plus faible que l'hypothèse (3.19). Cette section se divise comme il suit. La sous-section 3.3.1 présente le processus de contraste utilisé pour définir chacune des statistiques de test sur le $l^{\text{ème}}$ coefficient de Fourier. La sous-section 3.3.2 présente la procédure de test sur chacun des coefficients de Fourier. Enfin la sous-section 3.3.3 étudie la statistique de test pour l'homogénéité des fonctions de régression.

3.3.1 Estimation des paramètres de translation

Description du modèle. On considère dans cette section le modèle suivant,

$$Y_{i,j} = a_j^* f^*(t_i - \theta_j^*) + \sigma^* \epsilon_{i,j}, \quad j = 1 \dots J, \quad i = 1 \dots n. \quad (3.20)$$

f^* est une fonction inconnue, 2π -périodique et continue. Les paramètres $\theta^* = (\theta_j^*)_{j=1 \dots J}$ et $a^* = (a_j^*)_{j=1 \dots J}$ sont des paramètres de \mathbb{R}^J inconnus. Les points d'observation t_i sont équidistants sur $[0, 2\pi]$,

$$t_{i,j} = t_i = \frac{i-1}{n} 2\pi \in [0, 2\pi[, \quad j = 1 \dots J, \quad i = 1 \dots n.$$

Les termes de bruit $(\epsilon_{i,j})_{i,j}$ sont des variables gaussiennes centrées réduites indépendantes.

On munit le modèle des contraintes d'identifiabilité suivantes (similaires au jeu n°1 de la section 2.1.2) :

$$f \in \mathcal{F} \text{ et } (\theta^*, a^*) \in \mathcal{A}_1 = \{(\theta, a) \in [0, 2\pi[\times [a_{\min}, a_{\max}[, \text{ tel que } \theta_1 = 0, a_1 = 1\},$$

avec $0 < a_{\min} < 1 < a_{\max}$.

L'objet de cette sous-section est de définir un estimateur de θ^* sans s'occuper des paramètres d'amplitude a_j^* . Le paramètre de nuisance est donc la fonction vectorielle $(a_1^* f^*, \dots, a_J^* f^*)$, tandis que le paramètre fini-dimensionnel est (θ^*, σ^*) .

Définition du processus de contraste. Le processus de contraste pour l'estimation de θ^* est défini à partir de la transformée de Fourier discrète du modèle (3.13) (cf la section 2.2). Sa définition est similaire au processus de contraste (2.16) du chapitre 2 sauf que les coefficients rephasé $\tilde{c}_{j,l}$ sont les suivants,

$$\tilde{c}_{j,l}(\theta) = e^{il\theta_j} d_{j,l}, \quad j = 1 \dots J, \quad l = -(n-1)/2, \dots, (n-1)/2. \quad (3.21)$$

On note également \hat{c}_l leurs moyennes respectives,

$$\hat{c}_l(\theta) = \frac{1}{J} \sum_{j=1}^J \tilde{c}_{j,l}(\theta), \quad l = -(n-1)/2, \dots, (n-1)/2.$$

Ainsi, en choisissant convenablement une suite de poids (δ_l) de carré sommable, on considère le processus de contraste suivant :

$$W_n(\theta) = \frac{1}{J} \sum_{j=1}^J \sum_{l=-(n-1)/2}^{(n-1)/2} \delta_l^2 |\tilde{c}_{j,l}(\theta) - \hat{c}_l(\theta)|^2, \quad (3.22)$$

où $\theta \in \Theta = \{\theta \in \mathbb{R}^J, \theta_1 = 0\}$.

Consistance de l'estimateur. On montre que le processus de contraste (3.22) converge uniformément en probabilité sur Θ en probabilité vers la fonction de contraste :

$$\int_0^{2\pi} \frac{1}{J} \sum_{j=1}^J (\psi * f_j^*(t))^2 \frac{dt}{2\pi} - \int_0^{2\pi} \left(\frac{1}{J} \sum_{j=1}^J a_j^* \psi * f^*(t - \theta_j^* + \theta_j) \right)^2 \frac{dt}{2\pi}.$$

Cette fonction de contraste admet un unique minimum sous certaines conditions. Par ailleurs, au processus de contraste (2.16), on prouve les corollaires suivants sur la consistance et la normalité asymptotique de l'estimateur des translations induit comme le minimum du processus (3.22).

Corollaire 3.8. *Soit l'estimateur des translations défini comme*

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta} W_n(\theta),$$

où W_n désigne le processus (3.22). Alors, sous les mêmes hypothèses que le corollaire 3.2, $\hat{\theta}_n$ converge en probabilité vers θ^* .

Remarque 3.5. Pour que W_n aie asymptotiquement un unique minimum θ^* , on a vu au chapitre 2 qu'il était nécessaire qu'il existe deux entiers premiers entre eux l, k tels que $\delta_l \neq 0$, $\delta_k \neq 0$ et $c_l(f) \neq 0$, $c_k(f) \neq 0$. Dans la section suivante, on considère les

processus de contraste pour lesquels la suite de poids est égale (pour $l \neq 0$ un entier fixé) :

$$\delta_m = \begin{cases} 1 & \text{si } m = l \\ 0 & \text{sinon.} \end{cases}$$

Pour de telles suite de poids, l'estimateur est consistant à condition de réduire l'espace d'optimisation comme il suit :

$$\Theta = \{\theta \in [0, 2\pi/l]^J, \text{ tel que } \theta_1 = 0\}.$$

Corollaire 3.9. Soit $\hat{\theta}_n$ l'estimateur défini au corollaire 3.8. Sous les mêmes hypothèses que le corollaire 3.4, les fluctuations $n^{1/2}(\hat{\theta}_n - \theta^*)$ convergent en loi vers un vecteur gaussien centré de matrice de covariance :

$$\sigma^{*2} \frac{\sum_{l \in \mathbb{Z}} \delta_l^4 l^2 |c_l(f)|^2}{\left(\sum_{l \in \mathbb{Z}} \delta_l^2 l^2 |c_l(f)|^2\right)^2} (D_a^{-1} + \mathbb{I}_{J-1} \mathbb{I}_{J-1}^T),$$

avec $D_a = \text{diag}(a_2^*, \dots, a_J^*)$ et $\mathbb{I}_{J-1} = (1, \dots, 1)^T \in \mathbb{R}^{J-1}$.

Idée de la preuve. La preuve est similaire au théorème 2.2.

- i) *Convergence du gradient.* Comme dans la proposition 2.3, on montre que le gradient $\nabla W_n(\theta^*)$ converge en loi vers un vecteur gaussien. Plus précisément on a,

$$\sqrt{n} \nabla W_n(\theta^*) = \frac{2}{J} \sum_{l \in \mathbb{Z}} \delta_l^2 l |c_l(f)| (\langle g_l^j, X_l \rangle)_{j=2 \dots J} + o_{\mathbb{P}}(1),$$

où pour $j = 2, \dots, J$, $g_l^j = (g_l^{j,k})_{k=1 \dots 2J}$ est un vecteur de \mathbb{R}^{2J} ,

$$\sqrt{2} g_l^{j,k} = \begin{cases} \frac{-a_j^*}{J} \sin(l\theta_k^* - \gamma_l) & \text{si } j \in \{1 \dots J\} \setminus \{j\} \\ \left(\bar{a}^* - \frac{a_j^*}{J}\right) \sin(l\theta_j^* - \gamma_l) & \text{si } k = j \\ \left(\bar{a}^* - \frac{a_j^*}{J}\right) \cos(l\theta_j^* - \gamma_l) & \text{si } k = j + J \\ \frac{-a_j^*}{J} \cos(l\theta_{k-J}^* - \gamma_l) & \text{si } j \in \{J+1 \dots 2J\} \setminus \{J+j\} \end{cases}$$

et $\bar{a}^* = \sum_{j=1}^J a_j^*/J$, $X_l = (\xi_{1,l}^x, \dots, \xi_{J,l}^x, \xi_{1,l}^y, \dots, \xi_{J,l}^y)^T$, $\gamma_l = \arg(c_l(f))$. Ainsi, le gradient de W_n converge en loi vers un vecteur gaussien centré de matrice de covariance Γ_0 ,

$$\Gamma_0 = \frac{4}{J^2} \sum_{l \in \mathbb{Z}} \delta_l^4 l^2 |c_l(f)|^2 \left((\bar{a}^*) I_{J-1} - \frac{1}{J} a \mathbb{I}_{J-1}^T \right) \left((\bar{a}^*) I_{J-1} - \frac{1}{J} a \mathbb{I}_{J-1}^T \right)^T + \frac{1}{J^2} a a^T$$

où $a = (a_2^*, \dots, a_J^*)^T$.

- ii) *Convergence uniforme de la dérivée seconde.* Comme dans la proposition 2.4, pour tout suite $(\bar{\theta}_n)_n$ telle que $\|\bar{\theta}_n - \theta^*\| \leq \|\hat{\theta}_n - \theta^*\|$, la matrice hessienne $\nabla^2 M_n(\bar{\theta}_n)$ converge vers la matrice inversible $\frac{2}{J^2} \sum_{l \in \mathbb{Z}} \delta_l^2 l^2 |c_l(f)|^2 H$, avec

$$H = \bar{a}^* D_a - \frac{a a^T}{J}, \quad \text{et} \quad H^{-1} = \frac{D(a^*) + \mathbb{I}_{J-1} \mathbb{I}_{J-1}^T}{\bar{a}^*}.$$

□

On remarque que, pour tout $j = 1 \dots J$, $a_j^* = 1$, alors cette matrice est identique à celle calculée par Gamboa et al. (2005).

3.3.2 Test statistique sur un coefficient de Fourier

Le modèle général est le suivant,

$$Y_{i,j} = f_j^*(t_i) + \epsilon_{i,j}, \quad i = 1 \dots n, \quad j = 1 \dots J.$$

Dans la section 2.2.2, on a vu que si la série de Fourier des fonctions f_j^* était normalement convergente, il était raisonnable de considérer une version simplifiée de la DFT (3.13) du modèle en substituant les coefficients de Fourier discrets des fonctions de régression par les vrais coefficients de Fourier. (Cela revient en fait à considérer un modèle de bruit blanc). Par conséquent, le modèle étudié est :

$$d_{j,l} = c_l(f_j^*) + \xi_{j,l}/\sqrt{n}, \quad j = 1, \dots, J, \quad l = -(n-1)/2, \dots, (n-1)/2,$$

où $(\xi_{j,l})_{j,l}$ sont des variables complexes indépendantes gaussiennes centrées de variance 1 (plus de précision à la section 2.2)

Soit $l \in \mathbb{Z}$ un entier relatif fixé. Notre objectif dans cette section est de construire une procédure statistique visant à tester

$$\begin{aligned} H_0 : & \quad \text{pour tout } j = 2 \dots J, \quad |c_l(f_j^*)| = |c_l(f_1^*)| \\ & \quad \text{contre} \\ H_1 : & \quad \text{il existe } j \in \{2 \dots J\} \text{ tel que } |c_l(f_j^*)| \neq |c_l(f_1^*)|. \end{aligned}$$

Supposons que $\forall j = 1 \dots J \quad |c_l(f_j^*)| \neq 0$.

On définit la suite de poids suivante et les réels suivants

$$\begin{aligned} \delta_m &= \begin{cases} 1 & \text{si } m = l, \\ 0 & \text{sinon,} \end{cases} \quad m \in \mathbb{Z}, \\ a_j^* &= |c_l(f_j)/c_l(f_1)| \quad \text{et} \quad \theta_{j,l}^* = \arg(c_l(f_j)) - \arg(c_l(f_1)). \end{aligned} \tag{3.23}$$

Alors les $l^{\text{ème}}$ coefficient de Fourier des données est pour $j = 1 \dots J$,

$$d_{j,l} = a_j^* e^{-il\theta_j^*} c_l(f_1^*) + \xi_{j,l}/\sqrt{n}. \quad (3.24)$$

Écrit de cette façon, le modèle (3.24) est identique au modèle des coefficients de Fourier dans la section 3.3.1. Ceci correspond à la transformée de Fourier discrète du modèle étudié dans la sous-section précédente pour le $l^{\text{ème}}$ coefficient de Fourier et pour la forme commune f_1^* .

Notons $W_{n,l}(\theta)$ le processus de contraste (3.22) associé à la suite de poids (3.23).

On établit le théorème suivant.

Théorème 3.2. *Supposons que les hypothèses du corollaire 3.6 soient vérifiées. Soit $T_{n,l}$ la statistique de test définie à partir du processus de contraste $W_{n,l}$*

$$T_{n,l} = 2Jn \inf_{\theta \in \Theta} W_{n,l}(\theta)/\sigma_n^2.$$

Alors :

- **sous H_1** : $T_{n,l}$ tend vers l'infini avec n .
- **sous H_0** : $T_{n,l}$ converge en loi vers une variable du chi-deux avec $J - 1$ degrés de liberté $\chi^2(J - 1)$.

Remarquons que la loi limite de la statistique ne dépend ni de la forme commune f_1^* , ni de l'entier relatif l .

Supposons qu'il existe j tel que $|c_l(f_j)| = 0$.

En utilisant les mêmes arguments que le théorème 3.2, on a le corollaire suivant.

Corollaire 3.10. *Supposons qu'il existe j tel que $|c_l(f_j)| = 0$. Si H_1 est vraie alors, sous les hypothèses et les définitions du théorème 3.2, $T_{n,l}$ converge vers $+\infty$ avec n .*

Il reste à étudier le cas où H_0 est vraie. Dans ce cas pour tout $j = 1 \dots J$, on a que $c_l(f_j) = 0$, et la limite n'est plus une statistique du chi-deux. Par exemple pour $J = 2$, la statistique de test est égale à :

$$T_{n,l} = 2 \left(|\sqrt{n}\xi_{1,l}| - |\sqrt{n}\xi_{2,l}| \right)^2.$$

Ainsi $T_{n,l}$ est la différence de racines carrées de deux variables du chi-deux indépendantes à deux degrés de liberté. Après calculs, la densité de cette variable aléatoire est

$$f(x) = \left\{ e^{-1/4x} \left(\sqrt{x} e^{-1/4x} \sqrt{\pi} (\phi(x) - \phi(0)) (1/2x - 1) \right) \right\} / (4\sqrt{x}),$$

où ϕ est la fonction de répartition d'une loi normale centrée réduites.

J	risque du test	H_0		H_1	
		$n = 51$	$n = 201$	$n = 51$	$n = 201$
2	0.10	9.6	9.5	68.7	98.5
	0.05	5.1	5.7	60.7	97.7
	0.01	0.5	0.5	43.2	94.4
4	0.10	9.6	10.8	79.4	99.7
	0.05	4.6	5.5	71.0	99.5
	0.01	1.3	0.9	53.7	97.8

TAB. 3.2: Pourcentage de rejet de H_0 sous les hypothèses statistiques H_0 et H_1 , calculés à partir de 1000 tests avec $l = 1$.

Pour cette raison, on suppose par la suite que les coefficients de Fourier de la fonction de régression f_1^* soient non nul. Cela revient en fait à faire un traitement des données avant de mettre en oeuvre la procédure de test de la sous-section suivante. On doit sélectionner les indices $l \in \mathbb{Z}$ pour lesquels il existe au moins un coefficient $c_l(f_j^*)$, $j = 1 \dots J$, non nuls.

Simulations

On présente maintenant des simulations associées à la procédure de test décrite dans cette sous-section. Le tableau 3.2 illustre le théorème 3.2 pour $l = 1$. L'hypothèse nulle est "pour tout $j = 2, \dots, J$ $|c_1(f_j^*)| = |c_1(f_1^*)|$ ". Le tableau 3.2 donne le pourcentage de rejet sous l'hypothèse nulle H_0 et l'hypothèse alternative H_1 pour des données simulées à partir du modèle (3.24).

3.3.3 Agrégation des statistiques de test

Supposons pour simplifier que pour tout $l \in \mathbb{Z}$, $c_l(f_1^*) \neq 0$. D'après le théorème 3.2 et le corollaire 3.10, nous allons construire une statistique de test pour le test d'hypothèses suivant :

$$\begin{aligned}
 H_0 : \quad & \forall l \in \mathbb{Z}, \forall j \in \{2, \dots, J\} \quad |c_l(f_j^*)| = |c_l(f_1^*)|, \\
 & \text{contre} \\
 H_1 : \quad & \exists l \in \mathbb{Z}, \exists j \in \{2, \dots, J\} \text{ tel que } |c_l(f_j^*)| \neq |c_l(f_1^*)|.
 \end{aligned}$$

Comme on l'a indiqué en introduction, on ne teste pas exactement que les fonctions de régression soient translatées les unes des autres ($f_j^*(\cdot) = f_1^*(\cdot - \theta_j^*)$). Mais il s'agit d'une condition nécessaire (pas suffisante).

D'autres part, comme les fonctions de régression sont à valeurs réelles ($\forall l \in \mathbb{Z}$, $c_{-l}(f_j^*) = \overline{c_l(f_j^*)}$), il suffit de considérer seulement les indices l appartenant aux entiers naturels \mathbb{N} .

Soit m_n une suite d'entiers naturels tendant vers l'infini avec n . Comme la borne supérieur du support de la densité de probabilité du chi-deux est $+\infty$, et d'après le théorème 3.2, la statistique $T_n = \sup\{T_{n,l}, l = 0 \dots m_n\}$ converge presque sûrement sous H_0 vers $+\infty$ avec $n \rightarrow \infty$. On utilise alors la théorie des valeurs extrêmes (Resnick (1987) pour plus de détails) afin de normaliser la statistique T_n de telle sorte qu'elle converge en loi.

Théorème 3.3. *Supposons que les hypothèses du corollaire 3.6 soient vérifiées. Soit \mathbb{F} la fonction de répartition de la loi du chi deux à $J - 1$ degrés de liberté. Pour $n \in \mathbb{N}^*$, on pose*

$$b_n = \{1/(1 - \mathbb{F})\}^{-1}(m_n + 1) \\ \simeq 2 \log(m_n + 1) + \left\{ (J-3) \log(\log(m_n + 1)) - 2\Gamma\left(\frac{J-1}{2}\right) \right\} \left(1 + \frac{J-3}{2 \log(m_n + 1)}\right),$$

où Γ est la fonction gamma standard,

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt, \quad x \in \mathbb{R}_+^*.$$

Si la suite d'entiers $(m_n)_n$ est telle que

$$\sqrt{\log m_n} \, m_n^2 / \sqrt{n} = o(1), \quad (3.25)$$

alors,

- i) sous H_1 , $(T_n - b_n)/2$ tend vers $+\infty$,
- ii) sous H_0 , $(T_n - b_n)/2$ converge en loi vers un variable de Gumbel de fonction de répartition $x \in \mathbb{R} \rightarrow \exp(-e^{-x}) \in [0, 1]$.

Simulations

On présente maintenant des simulations associées à la procédure de test du théorème 3.3. Le tableau 3.3 donne le pourcentage de rejet sous l'hypothèse nulle H_0 et l'hypothèse alternative H_1 pour des données simulées. Chaque résultat est calculés à partir de 1000 procédures de test et pour $m_n = 50$. Les données sont générées de la même façon que dans la sous-section 3.2.3. Pour les données générées sous l'hypothèse alternative H_1 , on considère la fonction de régression f_2^* suivante :

- cas 1, $f_2(\cdot) = 21f^*(\cdot - \theta_2^*)$, il y a seulement une variation d'amplitude.

J	niveau de test	H_0		H_1	
		$n = 501(a)$	$n = 50001(a)$	cas 1 $n = 501(b1)$	cas 2 $n = 501(b2)$
2	0.10	4.8	6.6	71.1	85.1
	0.05	3.3	4.0	64.7	81.5
	0.01	1.3	1.6	51.2	71.6
3	0.10	9.8	13.5	83.2	93.9
	0.05	6.7	8.1	79.4	91.5
	0.01	3.0	2.8	68.8	84.1

TAB. 3.3: Pourcentages de rejet de H_0 sous les hypothèses H_0 et H_1 , calculés à partir de 1000 jeux de données et $m_n = 50$.

- cas 2, $f_2(\cdot) = 20f^*(\cdot - \theta_2^*) + 0.2$, un seul des coefficients de Fourier ($l = 0$) diffère en module.

Le défaut de ce test vient de l'utilisation de la théorie valeurs extrêmes. En effet, pour que la distribution de T_n soit proche de celle d'une loi de Gumbel, il faut que m_n soit assez grand. Or l'hypothèse (3.25) implique que la taille de l'échantillon n soit donc beaucoup plus grande. Ceci explique pourquoi, pour les simulations présentées ici, cette hypothèse n'est pas vérifiées. Malgré cela, on observe que les pourcentages de rejet ne sont pas trop éloignés de ceux espérés.

3.3.4 Preuves

Puisque chaque statistique de test est divisée par un estimateur consistant de la variance, on suppose pour simplifier que $\sigma^* = 1$.

Preuve du théorème 3.2

Soit $\hat{\theta}_{n,l} \in [0, 2\pi/l]$ tel que

$$\hat{\theta}_{n,l} \in [0, 2\pi/l] = \arg \min_{\theta \in \Theta} W_{n,l}(\theta).$$

La statistique $T_{n,l}$ est donc égale à $2JnW_{n,l}(\hat{\theta}_{n,l})$, et on a,

$$T_{n,l} = 2JnW_n^l(\theta_l^*) + 2Jn \left\{ W_n^l(\hat{\theta}_{n,l}) - W_n^l(\theta_l^*) \right\}.$$

D'après un développement de Taylor au voisinage de $\hat{\theta}_{n,l}$, il existe $\tilde{\theta}_{n,l}$ au voisinage de θ_l^* tel que :

$$T_{n,l} = 2JnW_{n,l}(\theta_l^*) - J[\sqrt{n}(\hat{\theta}_{n,l} - \theta_l^*)]^T \nabla^2 W_n(\tilde{\theta}_{n,l})[\sqrt{n}(\hat{\theta}_{n,l} - \theta_l^*)].$$

Un autre développement de Taylor du gradient de $W_{n,l}$ montre l'existence d'un point $\bar{\theta}_{n,l}$ au voisinage de θ_l^* tel que,

$$T_{n,l} = 2JnW_{n,l}(\theta_l^*) - J[\sqrt{n}\nabla W_{n,l}(\theta_l^*)]^T [\nabla^2 W_{n,l}(\bar{\theta}_{n,l})]^{-1} \nabla^2 W_{n,l}(\tilde{\theta}_{n,l}) [\nabla^2 W_{n,l}(\bar{\theta}_{n,l})]^{-1} [\sqrt{n}\nabla W_{n,l}(\theta_l^*)].$$

Or par définition de $W_{n,l}$ (cf. l'idée de la preuve du corollaire 3.9), son gradient se réécrit sous la forme :

$$n^{1/2}\nabla W_{n,l}(\theta_l^*) = \frac{2}{J}l|c_l(f)|G_l(X_l) + o_{\mathbb{P}}(1),$$

et sa matrice hessienne pour toute suite $(\bar{\theta}_n)_n$ dans un voisinage de θ_l^* vérifie,

$$\nabla^2 W_{n,l}(\bar{\theta}_n) = \frac{2}{J^2}l^2|c_l(f)|^2H + o_{\mathbb{P}}(1), \quad \text{où } H^{-1} = \frac{1}{\bar{a}^*} (D_a + \mathbb{I}_{J-1}\mathbb{I}_{J-1}^T).$$

On obtient alors que,

$$T_{n,l} = 2JnM_{n,l}(\theta_l^*) - 2J[G_l(X_l)]^T H_l^{-1} [G_l(X_l)] + o_{\mathbb{P}}(1).$$

En utilisant le lemme 3.1, on peut faire les conclusions suivantes.

Sous H_1 , on vient de voir que $T_{n,l} = 2JnW_{n,l}(\theta_l^*) + \mathcal{O}_{\mathbb{P}}(1)$. On déduit alors du lemme 3.1 que $T_{n,l}$ tend vers $+\infty$ avec n .

Sous H_0 , considérons le vecteur Gaussien V_l de loi $\mathcal{N}_J(0, I_J)$ suivant :

$$V_l = \sqrt{2} \left(\cos(l\theta_{1,l}^* - \gamma_l) \xi_{1,l}^x - \sin(l\theta_{1,l}^* - \gamma_l) \xi_{1,l}^y, \dots, \cos(l\theta_{J,l}^* - \gamma_l) \xi_{J,l}^x - \sin(l\theta_{J,l}^* - \gamma_l) \xi_{J,l}^y \right)^T.$$

La statistique $T_{n,l}$ se réécrit sous la forme :

$$(V_l)^T \left(I_J - \frac{1}{J} \mathbb{I}_J (\mathbb{I}_J)^T \right) V_l.$$

Afin de calculer la loi de la statistique, on diagonalise la matrice symétrique $I_J - \frac{1}{J} \mathbb{I}_J (\mathbb{I}_J)^T$ dans une base orthonormale. Comme les valeurs propres de cette matrice sont 0 et 1 de multiplicités respectives 1 et $J - 1$. Le théorème de Cochran nous permet de conclure. \square

Lemme 3.1. Avec les notations de la sous-section 3.3.2, on pose $\theta_l^* = (\theta_{j,l}^*)_{j=1\dots J}$.

Sous H_1 , $\lim_n 2JnW_{n,l}(\theta_l^*) = +\infty$.

Sous H_0 , en reprenant les mêmes notations que la preuve du corollaire 3.9 on a,

$$JnW_{n,l}(\theta_l^*) = 2 \sum_{j=2}^J (\langle g_l^j, X_l \rangle^2 + \langle h_l^j, X_l \rangle^2) + \sum_{j \neq k; j,k=2}^J (\langle g_l^j, X_l \rangle \langle g_l^k, X_l \rangle + \langle h_l^j, X_l \rangle \langle h_l^k, X_l \rangle), \quad (3.26)$$

où les vecteurs g_l^j et h_l^j sont définis dans la preuve qui suit.

Preuve du lemme 3.1. Soient p_l^j et q_l^j les vecteurs de \mathbb{R}^{2J} suivants,

$$(p_l^k(u))_j = \begin{cases} -\frac{1}{J} \cos u_j & \text{si } j \in \{1 \dots J\} \setminus \{k\}; \\ \frac{J-1}{J} \cos u_j & \text{si } j = k; \\ -\frac{J-1}{J} \sin u_j & \text{si } j = k + J; \\ \frac{1}{J} \sin u_j & \text{si } j \in \{J+1 \dots 2J\} \setminus \{J+k\} \end{cases}$$

$$(q_l^k(u))_j = \begin{cases} -\frac{1}{J} \sin u_j & \text{si } j \in \{1 \dots J\} \setminus \{k\}; \\ \frac{J-1}{J} \sin u_j & \text{si } j = k; \\ \frac{J-1}{J} \cos u_j & \text{si } j = k + J; \\ -\frac{1}{J} \cos u_j & \text{si } j \in \{J+1 \dots 2J\} \setminus \{J+k\} \end{cases}$$

où $u = (u_j) \in \mathbb{R}^J$. Posons $p_l^1 = -\sum_{j=2}^J p_l^j$ et $q_l^1 = -\sum_{j=2}^J q_l^j$. Après quelques calculs, on a,

$$2JnW_{n,l}(\theta_l^*) = 2n \sum_{j=1}^J |a_j^* - \bar{a}^*|^2 |c_l(f_1)|^2 \quad (3.27)$$

$$+ \sqrt{2n} \sum_{j=1}^J (a_j^* - \bar{a}^*) |c_l(f_1)| \langle p_l^j(l(\theta_l^*) - \gamma \mathbb{I}_J), X_l \rangle \quad (3.28)$$

$$+ \sum_{j=2}^J \langle p_l^j(l\theta_l^*), X_l \rangle^2 + \langle q_l^j(l\theta_l^*), X_l \rangle^2 + \left\langle \sum_{j=2}^J p_l^j(l\theta_l^*), X_l \right\rangle^2 + \left\langle \sum_{j=2}^J q_l^j(l\theta_l^*), X_l \right\rangle^2. \quad (3.29)$$

- i) Si H_1 est vraie. Il existe j tel que $a_j^* - \bar{a}^* \neq 0$, sinon cela signifierait que tous les a_j^* seraient égaux et l'hypothèse H_0 serait vérifiée. Par conséquent, le terme déterministe (3.27) converge vers $+\infty$ avec n . Par ailleurs, (3.28) est un vecteur gaussien centré dont la variance tend vers $+\infty$ avec n , alors que (3.29) est une suite tendue (bornée en probabilité). En conclusion, on a bien que :

$$\lim_n JnW_{n,l}(\theta_l^*) = +\infty.$$

ii) Si H_0 est vraie. Les termes (3.28) et (3.27) sont nuls. Posons $\gamma_l \in \Theta$ l'argument de $c_l(f_1^*) = |c_l(f_1^*)|e^{i\gamma_l}$. Ainsi, on a bien

$$g_l^j = -\sin \gamma_l p_l^j(l\theta_l^*) + \cos \gamma_l q_l^j(l\theta_l^*), \forall j = 1 \dots J.$$

Définissons les vecteurs $h_l^j \in \mathbb{R}^{2J}$ tels que :

$$h_l^j = \cos \gamma_l p_l^j(l\theta_l^*) + \sin \gamma_l q_l^j(l\theta_l^*), \forall j = 1 \dots J,$$

de telle sorte que les vecteurs $(h_l^j)_j$ sont orthogonaux aux vecteurs $(g_l^j)_j$:

$$\begin{bmatrix} q_l^j(l\theta_l^*) \\ p_l^j(l\theta_l^*) \end{bmatrix} = \begin{bmatrix} \cos \gamma_l & \sin \gamma_l \\ -\sin \gamma_l & \cos \gamma_l \end{bmatrix} \begin{bmatrix} g_l^j \\ h_l^j \end{bmatrix}.$$

En simplifiant, le lemme est prouvé. □

Preuve du théorème 3.3

Supposons que l'hypothèse H_0 soit vraie. Notons qu'afin d'assurer l'uniforme convergence du gradient $\nabla W_{n,l}(\theta_l^*)$ et de la matrice hessienne $\nabla^2 W_{n,l}(\theta_{n,l})$, une condition nécessaire est $m_n^2/\sqrt{n} = o(1)$. Soit T_l la variable aléatoire suivante

$$T_l = [G_l(X_l)]^T [H_l(\theta_l^*)]^{-1} [G_l(X_l)] + 2nW_{n,l}(\theta_l^*).$$

D'après le théorème 3.2 et l'hypothèse (3.25), $T_{n,l}$ converge en probabilité vers T_l . D'autre part, les variables $(T_l)_{l=1\dots m_n}$ sont des variables indépendantes identiquement distribuées de fonction de répartition \mathbb{F} (à cause de l'indépendance des variables $\xi_{j,l}$, cf. la section 2.2.1). Ainsi pour $x \in \mathbb{R}$ fixé, on a :

$$\left| \mathbb{P}\left(\sup_{l=1..m_n} T_{n,l} \leq 2x + b_n\right) - \exp(-e^{-x}) \right| \leq \left| \mathbb{P}\left(\sup_{l=1..m_n} T_l \leq 2x + b_n\right) - \exp(-e^{-x}) \right| \quad (3.30)$$

$$+ \left| \mathbb{P}\left(\sup_{l=1..m_n} T_{n,l} \leq 2x + b_n\right) - \mathbb{P}\left(\sup_{l=1..m_n} T_l \leq 2x + b_n\right) \right| \quad (3.31)$$

D'après la théorie des valeurs extrêmes (Resnick (1987)), la variable $\sup_{l=1..m_n} (T_l - b_n)/2$ converge en loi vers la distribution de Gumbel. Le premier terme (3.30) de l'inégalité converge donc vers zéro. Il reste alors à montrer que le second terme (3.31) de l'inégalité converge vers zéro.

Soient A_n, B_n les évènements respectifs $\left(\sup_{l=1..m_n} T_{n,l} \leq 2x + b_n\right)$ et $\left(\sup_{l=1..m_n} T_l \leq 2x + b_n\right)$. Nous utilisons l'inégalité suivante,

$$|\mathbb{P}(A_n) - \mathbb{P}(B_n)| \leq \mathbb{P}(A_n \setminus \{A_n \cap B_n\}) + \mathbb{P}(B_n \setminus \{A_n \cap B_n\}).$$

Soit Z_n la variable $\sup_{l=1..m_n} |T_l - T_{n,l}|$. D'après l'inégalité,

$$\sup_{l=1..m_n} T^l \leq \sup_{l=1..m_n} T_n^l + Z_n,$$

on obtient que,

$$\mathbb{P}(A_n \setminus \{A_n \cap B_n\}) \leq \mathbb{P}\left(\sup_{l=1..m_n} (T^l - b_n)/2 + Z_n/2 \leq x\right) - \mathbb{P}\left(\sup_{l=1..m_n} T^l \leq 2x + b_n\right) \quad (3.32)$$

Puisque le lemme 3.2 atteste que sous les hypothèses du théorème, la variable Z_n converge vers zéro en probabilité, $\mathbb{P}(A_n \setminus \{A_n \cap B_n\})$ converge vers zéro avec n . De la même façon, $\mathbb{P}(B_n \setminus \{A_n \cap B_n\})$ converge vers 0. Le théorème est montré sous H_0 .

Supposons que l'hypothèse H_1 soit vraie. D'après la preuve du théorème 3.2, il existe une constante positive $c > 0$ telle que $T_n = cn + \mathcal{O}_{\mathbb{P}}(\sqrt{n})$. Comme $b_n = \mathcal{O}(\log(m_n))$, on déduit que $(T_n - b_n)/2$ tends vers $+\infty$. \square

Lemme 3.2. *Sous les hypothèses du théorème 3.3, on a :*

$$\sup\{|T_{n,l} - T_l|, l = 0 \dots m_n\} = o_{\mathbb{P}}(1).$$

La preuve du lemme nécessite l'introduction de notations décrites ci dessous.

Soit $C = (c_{k,l})_{k,l}$ une matrice carrée de \mathbb{R}^{J-1} , et $X = (x_1, \dots, x_{J-1}) \in \mathbb{R}^{J-1}$. On considère l'espace \mathbb{R}^{J-1} muni de la norme

$$\|X\|_{\infty} = \sup_{j=1 \dots J-1} |x_j|.$$

On associe à cette norme sur les vecteurs, la norme sur les opérateurs (matrices) de \mathbb{R}^{J-1} dans \mathbb{R}^{J-1} suivante,

$$\|C\|_{\infty} = \sup_{\|X\|_{\infty} < 1} \|CX\|_{\infty} = \sup_{k=1 \dots J-1} \sum_{j=1}^{J-1} |c_{k,j}|.$$

Pour $l = 0, \dots, (n-1)/2$, on note par X_l le vecteur aléatoire suivant,

$$X_l = (\xi_{1,l}^x, \dots, \xi_{J,l}^x, \xi_{1,l}^y, \dots, \xi_{J,l}^y)$$

(où $\xi_{j,l}^x = \Re \xi_{j,l}$ et $\xi_{j,l}^y = \Im \xi_{j,l}$ pour $j = 1, \dots, J$). Ainsi, d'après le lemme 2.1, X_l est un vecteur gaussien centré $\mathcal{N}_{2J}(0, I_{2J}/2)$ pour tout $l > 0$. D'après le chapitre 2 de Resnick (1987), le théorème des valeurs extrêmes affirme que,

$$\begin{aligned} \sqrt{\log m_n} \sup_{l=1..m_n} \|X_l\|_{\infty} - \log m_n &= O_{\mathbb{P}}(1) \\ \frac{1}{\sqrt{\log m_n}} \sup_{l=1..m_n} \|X_l\|_{\infty} &\xrightarrow{\mathbb{P}} 1. \end{aligned} \quad (3.33)$$

Preuve du lemme 3.2. D'après le théorème 3.2, on a,

$$T_{n,l} = JnW_{n,l}(\theta_l^*) - [\sqrt{n}\nabla W_{n,l}(\theta_l^*)]^T [\nabla^2 W_{n,l}(\bar{\theta}_{n,l})]^{-1} \nabla^2 W_{n,l}(\tilde{\theta}_{n,l}) [\nabla^2 W_{n,l}(\bar{\theta}_{n,l})]^{-1} [\sqrt{n}\nabla W_{n,l}(\theta_l^*)].$$

De plus, en reprenant l'idée de la preuve du corollaire 3.9 et la preuve des propositions 2.3 et 2.4, on peut écrire :

$$\begin{aligned} \sqrt{n}\nabla W_{n,l}(\theta^*) &= lG_l(X_l) + \frac{l}{\sqrt{n}}B_l(X_l) \\ \nabla^2 W_{n,l}(\theta) &= l^2H_l(\theta) + \frac{l^2}{\sqrt{n}}C_l(X_l, \theta) + \frac{l^2}{n}D_l(X_l, \theta), \end{aligned}$$

où $G_l(X_l)$, $B_l(X_l)$ sont des vecteurs aléatoires de \mathbb{R}^{J-1} dont les coordonnées sont respectivement des formes linéaires et quadratiques de X_l . $C_l(X_l, \theta)$, $D_l(X_l, \theta)$ sont des matrices carrées aléatoires de \mathbb{R}^{J-1} dont les coordonnées sont respectivement des formes linéaires et quadratiques de X_l . Par la continuité de G_l , B_l , C_l et D_l en la variable X_l , il existe une constante positive $\kappa > 0$ telle que :

$$\begin{aligned} \|G_l(X_l)\|_\infty &\leq \kappa \|X_l\|_\infty, & \|B_l(X_l)\|_\infty &\leq \kappa \|X_l\|_\infty^2, \\ \forall \theta \in A \quad \|C_l(X_l, \theta)\|_\infty &\leq \kappa \|X_l\|_\infty, & \|D_l(X_l, \theta)\|_\infty &\leq \kappa \|X_l\|_\infty^2. \end{aligned}$$

Soit $((H_l(\theta))_{k,j})_{2 \leq k,j \leq J}$ les éléments de la matrice carrée de taille $J-1$ déterministe $H_l(\theta)$,

$$(H_l(\theta))_{k,j} = \begin{cases} -|c_l(f)|^2 \cos \{l(\theta_k - \theta_k^* + \theta_j^* - \theta_j)\} & \text{si } k \neq j \\ |c_l(f)|^2 \sum_{j_0=1}^J \cos \{l(\theta_k - \theta_k^* + \theta_{j_0}^* - \theta_{j_0})\} & \text{sinon.} \end{cases}$$

Tout d'abord, montrons que,

$$\sup_{l=1..m_n} \|H_l(\bar{\theta}_{n,l})^{-1} H_l(\tilde{\theta}_{n,l}) H_l(\bar{\theta}_{n,l})^{-1} - H_l(\theta_l^*)^{-1}\|_\infty = o_{\mathbb{P}}(1). \quad (3.34)$$

En effet, on remarque que,

$$\begin{aligned} \sup_{l=1..m_n} \left\{ \|l(\theta_l^* - \theta)\|_\infty, \theta \in \{\tilde{\theta}_{n,l}, \bar{\theta}_{n,l}\} \right\} &\leq \sup_{l=1..m_n} \|l(\theta_l^* - \hat{\theta}_{n,l})\|_\infty \\ &\leq \frac{m_n^2}{\sqrt{n}} \sup_{l=1..m_n} \|G_l(X_l) + \frac{1}{\sqrt{n}}B_l(X_l)\|_\infty \\ &\leq \kappa \frac{\sqrt{\log m_n} m_n^2}{\sqrt{n}} \left(\frac{\sup_{l=1..m_n} \|X_l\|_\infty}{\sqrt{\log m_n}} + \frac{\sqrt{\log m_n}}{\sqrt{n}} \frac{\sup_{l=1..m_n} \|X_l\|_\infty^2}{\log m_n} \right). \end{aligned}$$

Du théorème des valeurs extrêmes (3.33) et de l'hypothèse (3.25), on déduit que,

$$\sup_{l=1..m_n} \left\{ \|l(\theta_l^* - \theta)\|_\infty, \theta \in \{\tilde{\theta}_{n,l}, \bar{\theta}_{n,l}\} \right\} = o_{\mathbb{P}}(1).$$

Comme $c_l(f) = \mathcal{O}(1)$, le théorème de l'application continue et l'uniforme continuité du cosinus impliquent que,

$$\sup_{l=1..m_n} \left\{ \|H_l(\theta) - H_l(\theta_l^*)\|_\infty, \theta \in \{\tilde{\theta}_{n,l}, \bar{\theta}_{n,l}\} \right\} = o_{\mathbb{P}}(1).$$

Or l'ensemble $GL_{J-1}(\mathbb{R})$ des matrices inversibles de \mathbb{R}^{J-1} est un ouvert, donc l'application $H \in GL_{J-1}(\mathbb{R}) \rightarrow H^{-1} \in GL_{J-1}(\mathbb{R})$ est continue. Par conséquent, l'équation (3.34) est établie.

D'autre part, si T_l est la variable aléatoire $[G_l(X_l)]^T [H_l(\theta_l^*)]^{-1} [G_l(X_l)] + 2nM_n^l(\theta_l^*)$, après quelques calculs et d'après (3.34), on obtient :

$$\begin{aligned} T_{n,l} = T_l + o_{\mathbb{P}}(1) &+ \sum_{k=1}^6 \left(G_l(X_l) + \frac{1}{\sqrt{n}} B_l(X_l) \right)^T \frac{P_{l,k}(X_l, \bar{\theta}_n^l, \tilde{\theta}_n^l)}{\sqrt{n}^k} \left(G_l(X_l) + \frac{1}{\sqrt{n}} B_l(X_l) \right) \\ &+ B_l(X_l)^T \frac{H_l(\theta_l^*)}{n} B_l(X_l) + B_l(X_l)^T \frac{2H_l(\theta_l^*)}{\sqrt{n}} G_l(X_l), \end{aligned}$$

où $P_{l,k}(X_l, \bar{\theta}, \tilde{\theta})$ est une matrice carrée de \mathbb{R}^{J-1} dont les coefficients sont des polynômes homogènes de degré k en X_l et tels que :

$$\forall (\bar{\theta}, \tilde{\theta}) \in A^2, \quad \|P_{l,k}(X_l, \bar{\theta}, \tilde{\theta})\|_\infty \leq \kappa \|X_l\|_\infty^k.$$

Par conséquent, on a :

$$\begin{aligned} \sup_{l=1..m_n} |T_{n,l} - T_l| &\leq \kappa^3 \sum_{k=1}^6 \frac{\sup_{l=1..m_n} \|X_l\|_\infty^{k+2}}{\sqrt{n}^k} + 2 \frac{\sup_{l=1..m_n} \|X_l\|_\infty^{k+3}}{\sqrt{n}^{k+1}} + \frac{\sup_{l=1..m_n} \|X_l\|_\infty^{k+4}}{\sqrt{n}^{k+2}} \\ &+ \kappa^3 \frac{\sup_{l=1..m_n} \|X_l\|_\infty^4}{n} + 2\kappa^3 \frac{\sup_{l=1..m_n} \|X_l\|_\infty^3}{\sqrt{n}} + o_{\mathbb{P}}(1). \end{aligned}$$

Et donc d'après (3.33) et l'hypothèse (3.25), le lemme est prouvé. \square

Chapitre 4

Estimation des rotations, des translations et des homothéties entre des images bruitées^{*,†}

Sommaire

4.1	Un modèle semi-paramétrique pour l'alignement affine	120
4.2	Transformées de Fourier et invariance par $\mathcal{R.S.T.}$	125
4.3	Estimation des paramètres	127
4.4	Simulations	137
4.5	Preuves	140

Résumé : L'objet d'étude de ce chapitre est l'alignement affine d'un ensemble fini d'images bruitées. Tout d'abord, nous définissons un modèle statistique adapté pour l'alignement affine. À partir de ce modèle, des critères basés sur les transformées de type Fourier sont proposés pour estimer les paramètres de rotation, d'échelle et de translation afin d'aligner ces images. Ces critères constituent une procédure statistique en deux étapes qui ne requiert pas l'utilisation d'une image de référence sur laquelle on alignerait les images. Notre approche est celle de la M -estimation. La consistance des estimateurs induits par ces critères est prouvée. Un exemple réel et des simulations illustrent les performances numériques de notre procédure.

Mots clés : Alignement d'image, transformation affine, estimation semi-paramétrique, modèle de bruit blanc, transformée de Fourier, transformée de Fourier Mellin.

^{*}**Remerciements :** Nous sommes reconnaissant à Stéphane Derrode de nous avoir fourni ses bases d'images de papillons

[†]Ce chapitre est un travail en collaboration avec Fabrice Gamboa et Jérémie Bigot

Introduction

Dans le traitement de l'image, une tâche fondamentale préalable est l'alignement de deux ou plusieurs images prises, par exemple, à différentes heures, par différents appareils ou de différents points de vue. Généralement, pour comparer des objets semblables, il est nécessaire de trouver une référence commune pour les représenter. À partir de cette référence commune, on cherche à aligner les images, c'est-à-dire à calculer la fonction qui déforme l'image de celle de référence et qui permet par la suite d'aligner les images. Le problème de l'alignement d'images a des applications dans des domaines assez variés, comme le traitement d'images satellites (Richards (2005) chapitre 2), ou encore en imagerie médicale (Hajnal et al. (2001)). La fonction de déformation peut être paramétrique ou non paramétrique, mais dans tous les cas elle doit être bijective. Au fil des années, une large gamme de techniques d'alignement a été développée pour différents types de données et de problèmes. On peut se reporter aux articles de Brown (1992) et Glasbey (1998) pour une revue détaillée de ces techniques, ainsi que les articles de Glasbey et Mardia (2001) et Markussen (2004) pour une analyse statistique des problèmes de déformation d'images.

Dans ce chapitre, nous nous focalisons sur les déformations affines qui conservent les rapports des distances et les angles orientés : les similitudes directes. Les similitudes directes sont l'identité du plan, les translations, les rotations, les homothéties, ou une composée des applications précédemment citées. Lorsque deux images diffèrent seulement d'une translation, la technique de corrélation de phase (Kuglin et Hines (1975)) est une approche sur le domaine des fréquences de Fourier souvent employée pour déterminer le vecteur de translation. Cette technique est basée sur la propriété de diagonalisation de l'opérateur de translation par la base de Fourier. D'autre part, lorsque les images diffèrent seulement par une rotation et une homothétie, l'expression en coordonnées log-polaires des images permet de revenir au modèle où les images diffèrent uniquement par une translation. Ainsi deux techniques sont envisageables pour estimer les paramètres relatifs aux opérateurs de rotation et d'homothétie : soit l'emploi de la corrélation de phase à partir des coordonnées log-polaires, soit l'emploi de la transformée de Fourier-Mellin qui est un type de transformée de Fourier définie sur les coordonnées polaires (Derrode et Ghorbel (2004)). En combinant ces deux techniques, diverses méthodes pour aligner des images ont été construites. Ces méthodes sont bien souvent des procédés en deux voire trois étapes. Généralement, la première étape considère le module de la transformée de Fourier afin d'estimer l'angle de rotation et le rapport d'homothétie. L'estimation de ces paramètres utilise soit la technique de corrélation de phase (Reddy et Chatterji (1996), De Castro et Morandi

(1987) et McGuire (1998)), ou soit la fonction de différence angulaire (Keller et al. (2005a,b)) qui est construite à partir de la transformée de Fourier pseudo-polaire (qui consiste principalement à calculer la transformée de Fourier sur une grille de points spécifiques à mi-chemin entre la grille cartésienne et polaire). La seconde étape est alors d'estimer les paramètres de translation à partir des transformées de Fourier normalisées.

C'est un de ces procédés en deux étapes, ayant pour but d'aligner un ensemble d'images bruitées qui diffèrent les unes des autres par une similitude et dont la forme générale est inconnue, que nous proposons d'analyser d'un point de vue statistique. Notre contribution est double. Premièrement, nous donnons de nouveaux critères pour aligner simultanément un ensemble d'images qui n'exigent pas l'utilisation d'une image de référence sur laquelle les autres images seraient alignés. Deuxièmement, nous prouvons la consistance de cette procédure d'un point de vue asymptotique quand les images sont observées avec un bruit blanc. La technique est basée sur la définition de fonctions de contrastes appropriées. La preuve de la convergence suit les lignes classiques de celle des M -estimateurs (voir le chapitre 1).

Le plan de ce chapitre est le suivant. La section 4.1 présente le modèle semi-paramétrique pour le problème d'alignement où les déformations sont des similitudes directes. Nous présentons quelques notations et quelques outils mathématiques sur ce modèle. En particulier, l'usage d'un modèle de type bruit blanc fournit un cadre théorique intéressant pour l'étude asymptotique que nous voulons mener. Puis nous discutons des contraintes d'identifiabilité posées sur ce modèle. La section 4.2 rappelle les propriétés d'invariance de la transformée de Fourier pour la translation, et de la transformée de Fourier-Mellin pour la rotation et l'homothétie. Nous rappelons également la définition de la transformée de Fourier Mellin analytique proposée par (Derrode et Ghorbel (2004)). Puis dans la section 4.3, les processus de contraste, pour l'estimation des paramètres caractérisant les similitudes associées aux images, sont définis. Dans une première étape, nous estimons les paramètres de rotation et d'échelle, et dans un deuxième temps, nous estimons les paramètres de translation. La consistance de ces procédures est étudiée d'un point de vue asymptotique, les preuves peuvent être trouvées dans la section 4.5. Enfin, la section 4.4 illustre les performances numériques de la méthode d'estimation sur des données simulées et réelles.

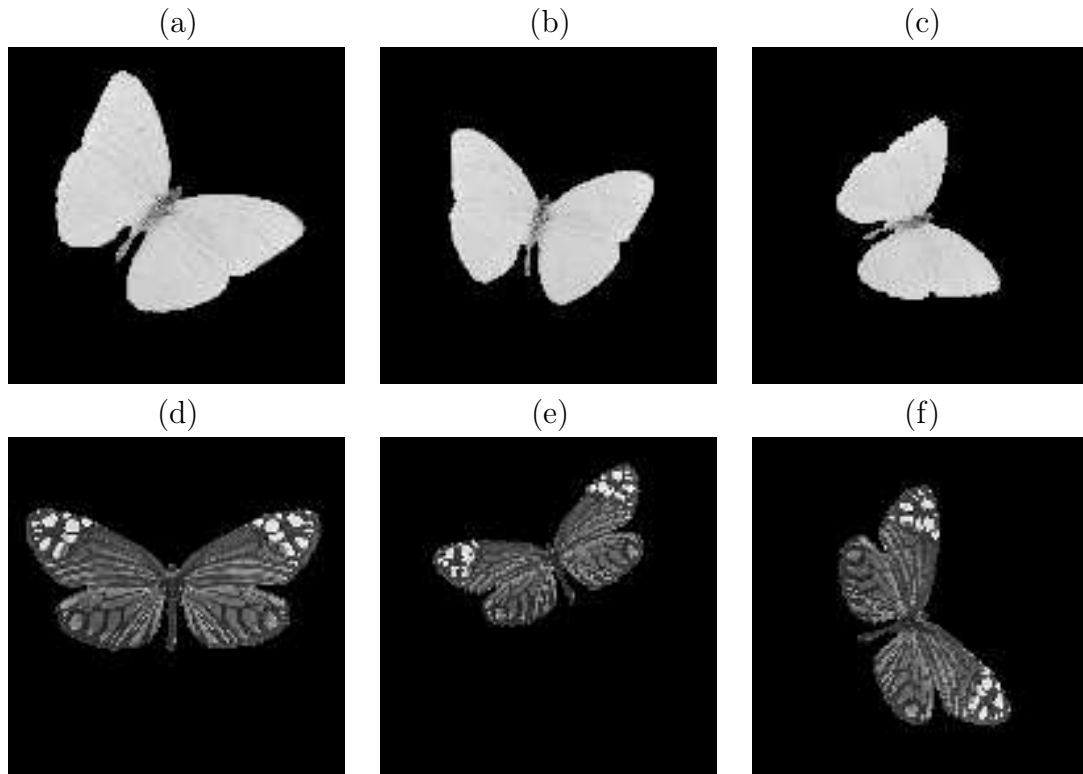


FIG. 4.1: Les images (a),(b) et (c) représentent le même papillon observé avec des tailles, des orientations et des emplacements différents. Les images (d),(e) et (f) représentent un autre papillon de forme différente, avec des tailles, des orientations et des emplacements divers.

4.1 Un modèle semi-paramétrique pour l'alignement affine

Pour motiver le modèle semi-paramétrique décrit par la suite, on peut considérer le problème d'alignement suivant. La figure 4.1 présente des images (probablement bruitées) de deux espèces de papillons ayant des formes, des orientations et des emplacements différents. La solution d'un problème d'alignement d'images consiste à récupérer les paramètres d'échelle, de rotation et de translation pour aligner les images associées à la même espèce de papillon (c'est-à-dire soit à la première ligne ou soit la seconde ligne de la figure 4.1). On dit que les images sont alignées s'il n'y a aucune différence de taille, d'orientation et d'emplacement entre les images.

4.1.1 Description du modèle

Géométrie semblable

Notre but est d'aligner un ensemble fini d'images de dimension deux (2D) perturbées par un bruit gaussien en utilisant des déformations affines qui conservent les angles orientés et les rapports de distances. De telles transformations sont appelées des similitudes directes. Elles sont la composée de trois transformations de base : l'homothétie, la rotation et la translation. L'ensemble des similitudes directes constitue donc une famille paramétrique $\phi_{a,\theta,b} : \mathbb{R}^2 \mapsto \mathbb{R}^2$, de déformations (bijectives) du plan, où $1/a \in \mathbb{R}_+$ est le rapport de l'homothétie, $\theta \in \mathbb{T}$ (où \mathbb{T} désigne le tore $\mathbb{R}/(2\pi\mathbb{Z})$) est l'angle de rotation, et $b \in \mathbb{R}^2$ est le vecteur de translation. Plus précisément, la fonction $\phi_{a,\theta,b}$ est définie comme :

$$\phi_{a,\theta,b}(x) = \frac{1}{a}R_\theta(x - b), \quad \text{pour tout } x \in \mathbb{R}^2,$$

où R_θ est la matrice 2×2 de rotation,

$$R_\theta = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}.$$

Modélisation des images

Les images sont généralement observées sur un domaine borné du plan. Elles peuvent donc être modélisées par des fonctions $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ à valeurs réelles ayant un support inclus dans un compact \mathcal{D} de \mathbb{R}^2 (si on choisit de coder par zéro un pixel noir). On note par $\mathbb{L}^p(\mathcal{D})$ ($p \in \mathbb{N}$) l'espace des fonctions p -intégrable sur \mathbb{R}^2 dont le support est inclus dans le compact \mathcal{D} . Cet espace est muni de la norme $\|g\|_{\mathbb{L}^p(\mathcal{D})} = (\int_{\mathcal{D}} |g(x)|^p dx / 2\pi)^{1/p}$. Par conséquent, $\mathbb{L}^2(\mathcal{D})$ est un sous-espace de $\mathbb{L}^1(\mathcal{D})$. Ainsi pour toute fonction $g \in \mathbb{L}^2(\mathcal{D})$, sa déformation par une similitude directe $\phi_{a,\theta,b}$ est la fonction $g \circ \phi_{a,\theta,b}$ de $L^2(\mathbb{R}^2)$ à support compact. Pour simplifier l'étude, on fixe \mathcal{D} comme le carré de côté $d > 0$ centré à l'origine.

Le modèle des observations

Pour un ensemble de J images bruitées, on considère le modèle de bruit blanc suivant :

$$dY_j(x) = f(\phi_{a_j^*, \theta_j^*, b_j^*}(x))dx + \epsilon dW_j(x), \quad x \in \mathcal{D}, \quad j = 1, \dots, J \quad (4.1)$$

où $W_j, j = 1, \dots, J$ sont des mouvements brownien standard sur \mathcal{D} , et ϵ est un paramètre de niveau de bruit inconnu. La fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ est une image inconnue

appartenant à l'espace $\mathbb{L}^2(\mathcal{D})$. Les paramètres $a_j^*, \theta_j^*, b_j^*, j = 1, \dots, J$ sont respectivement les paramètres d'échelle, de rotation et de translation qui caractérisent l'image j et que l'on souhaite estimer. Les J images sont donc modélisées par les fonctions à support compact $f_j^* = f \circ \phi_{a_j^*, \theta_j^*, b_j^*}, j = 1, \dots, J$ de $\mathbb{L}^2(\mathbb{R}^2)$. Dans la pratique, les images sont toujours observées dans un domaine de \mathbb{R}^2 que nous supposons ici inclus dans \mathcal{D} . Ainsi la forme commune des images f est supposée définie sur un sous-domaine Ω de \mathcal{D} (par exemple $[-d/4, d/4]^2$) de telle sorte que chacune des images f_j^* soient entièrement observées sur \mathcal{D} . Afin de définir proprement les images f_j^* , on pose les hypothèses suivantes :

- A1** l'image f est nulle en dehors de Ω et vérifie certaines conditions de régularité définies plus tard.
- A2** les paramètres d'échelle, de rotation et de translation appartiennent à l'ensemble suivant,

$$(a_j^*, b_j^*, \theta_j^*) \in [a_{\min}, a_{\max}] \times [-b_{\max}, b_{\max}]^2 \times [0, 2\pi[, \text{ pour tout } j = 1, \dots, J, \quad (4.2)$$

où a_{\min}, a_{\max} et b_{\max} sont des constantes strictement positives définies par l'utilisateur, ces constantes reflètent la connaissance préalable sur les paramètres au vu des données. Notons également que l'intervalle semi-ouvert $[0, 2\pi[$ est compris comme la classe des représentants du tore $\mathbb{T} = \mathbb{R}/(2\pi\mathbb{Z})$.

- A3** les paramètres d'échelle, de translation et les angles de rotation sont tels que pour tout $1 \leq j \leq J$ la déformation du sous-domaine Ω par la transformation $\phi_{a_j^*, \theta_j^*, b_j^*}$ est incluse dans \mathcal{D} , c'est-à-dire pour tout $x \in \Omega$ et pour tout $j = 1, \dots, J$

$$\phi_{a_j^*, b_j^*, \theta_j^*}(x) \in \mathcal{D}.$$

Avec les hypothèses (**A1-A3**), toutes les images f_j^* appartiennent à $\mathbb{L}^2(\mathcal{D})$. Par conséquent, ces J images sont J versions déformées par une similitude de la même image f observée sur un fond noir (si on choisit de coder par zéro un pixel de couleur noir). Ce modèle correspond à l'exemple réel énoncé au début de la section (voir la première rangée et la seconde rangée d'images de la figure 4.1).

On analyse les propriétés asymptotiques des M -estimateurs pour les paramètres a_j^*, b_j^* et θ_j^* , c'est-à-dire quand le niveau de bruit ϵ tend vers zéro dans le modèle (4.1). Le modèle de bruit blanc (4.1) est un modèle continu qui a fait ses preuves en tant qu'outil théorique pour l'étude de problème de régression non-paramétrique pour des images 2D. Au premier abord, le modèle (4.1) peut sembler inapproprié, puisque dans la pratique les images sont discrétisées sur une grille régulière. Cependant, plusieurs auteurs ont prouvé que les résultats asymptotiques établis pour des modèles

de bruit blanc étaient comparables à ceux obtenus pour des modèles de régression non-paramétriques où les données sont observées en des points équidistants. Nous renvoyons aux travaux de [Brown et Low \(1996\)](#) et [Donoho et Johnstone \(1999\)](#) pour une étude détaillée des relations entre les modèles continus et les modèles de régression discrets. Rappelons que, par définition, observer dans un modèle de bruit blanc signifie que le statisticien a accès à toutes les intégrales stochastiques $\int_{\mathcal{D}} g(x) dY_j(x)$ pour tout élément g de $\mathbb{L}^2(\mathcal{D})$ et pour tout Y_j solution de l'équation (4.1) :

$$\int_{\mathcal{D}} g(x) dY_j(x) = \int_{\mathcal{D}} g(x) f_j^*(x) dx + \epsilon \int_{\mathcal{D}} g(x) dW_j(x).$$

Cette dernière équation signifie que $\int_{\mathcal{D}} g(x) dY_j(x)$ est une variable normale de moyenne $\int_{\mathcal{D}} g(x) f_j^*(x) dx$ et de variance $\epsilon^2 \int_{\mathcal{D}} g(x)^2 dx$.

Dans le contexte des problèmes inverses pour des images 2D, [Candès et Donoho \(2002\)](#) ont obtenu des résultats théoriques en modélisant des données topographiques bruitées avec des modèles de bruit blanc. L'analyse de [Candès et Donoho \(2002\)](#) fournit des estimateurs qui peuvent être numériquement calculés à partir de données issues d'un modèle de régression discret, et qui permettent d'obtenir des résultats très satisfaisants pour des images réelles. Notre but est d'obtenir pour le modèle de bruit blanc (4.1) des résultats similaires à ceux obtenus par [Candès et Donoho \(2002\)](#). De plus, pour le modèle où les images 2D sont observées sur une grille cartésienne discrète de points, l'interpolation de l'image pour calculer sa déformation par une transformation affine complique de manière significative l'analyse asymptotique des problèmes d'alignement présentés dans ce chapitre. Par conséquent, nous préférons éviter cette approche afin de se concentrer sur les propriétés asymptotiques des estimateurs plutôt que d'étudier le biais introduit par la discrétisation, bien qu'il soit discuté dans la section 4.4.

4.1.2 Identifiabilité du modèle

Soit \mathcal{A} l'espace des paramètres $[a_{min}, a_{max}]^J \times [0, 2\pi[^J \times [-b_{max}, b_{max}]^{2J}$. Si l'on remplace $\alpha^* = (a_1^*, \dots, a_J^*, \theta_1^*, \dots, \theta_J^*, b_1^*, \dots, b_J^*) \in \mathcal{A}$ par le vecteur,

$$\alpha = \begin{pmatrix} a_1 \\ \vdots \\ a_J \\ b_1 \\ \vdots \\ b_J \\ \theta_1 \\ \vdots \\ \theta_J \end{pmatrix} = \begin{pmatrix} a_1^* a_0 \\ \vdots \\ a_J^* a_0 \\ b_1^* + a_1 A_{-\theta_1} b_0 \\ \vdots \\ b_J^* + a_J A_{-\theta_J} b_0 \\ \theta_1^* + \theta_0 \\ \vdots \\ \theta_J^* + \theta_0 \end{pmatrix},$$

avec $a_0 \in]0, +\infty[$, $\theta_0 \in [0, 2\pi[$, $b_0 \in \mathbb{R}^2$, alors le modèle (4.1) est inchangé en substituant à la fonction f la fonction $f \circ \phi_{a_0, \theta_0, b_0}$. Ce modèle n'est donc pas identifiable. Des contraintes d'identifiabilité sont ajoutées au modèle (4.1) visant à réduire l'espace des paramètres \mathcal{A} ,

$$a_1^* = 1, \theta_1^* = 0, b_1^* = (0, 0), \quad (4.3)$$

Notons qu'avec ces conditions d'identifiabilité, on suppose implicitement que le sous-domaine Ω et les constantes a_{min} , a_{max} sont telles que $a_{min} < 1 < a_{max}$. Ces contraintes garantissent la convergence de nos estimateurs (voir les théorèmes 4.1 et 4.3). L'espace des paramètres dorénavant considéré est $\mathcal{A}_1 \times \mathcal{A}_2 \subset \mathcal{A}$, où l'ensemble \mathcal{A}_2 est l'espace des paramètres de translation,

$$\mathcal{A}_2 = \{(b_1, \dots, b_J) \in [-b_{max}, b_{max}]^{2J}, \text{ tel que } b_1 = (0, 0)\},$$

et l'ensemble \mathcal{A}_1 est l'espace des paramètres d'échelle et des angles de rotations,

$$\mathcal{A}_1 = \{(a_1, \dots, a_J, \theta_1, \dots, \theta_J) \in [a_{min}, a_{max}]^J \times [0, 2\pi[^J, \text{ tel que } a_1 = 1, \theta_1 = 0\}.$$

Ces contraintes d'identifiabilité signifient que les paramètres d'échelle, de rotation et de translation sont calculés à partir de la première image f_1^* . Cependant, comme nous le verrons, les images ne sont pas recalées à partir de l'image f_1^* , mais à partir de la moyenne des images recalées.

4.2 Transformées de Fourier et invariance par translation, rotation et homothétie

4.2.1 Transformée de Fourier standard

Soit $\omega \in \mathbb{R}^2$, la transformée de Fourier d'une fonction $g \in L^1(\mathbb{R}^2)$,

$$\hat{g}(\omega) = \int_{\mathbb{R}^2} g(x) e^{-i\omega \cdot x} \frac{dx}{2\pi},$$

où $\omega \cdot x$ est le produit scalaire entre ω et x . Lorsque g appartient à $\mathbb{L}^1(\mathbb{R}^2) \cap \mathbb{L}^2(\mathbb{R}^2)$, alors \hat{g} appartient aussi à $\mathbb{L}^2(\mathbb{R}^2)$. Il est alors bien connu que la transformée de Fourier se prolonge en une isométrie de $\mathbb{L}^2(\mathbb{R}^2)$ dans $\mathbb{L}^2(\mathbb{R}^2)$, et que pour toutes fonctions f et g de $\mathbb{L}^1(\mathbb{R}^2) \cap \mathbb{L}^2(\mathbb{R}^2)$, on a,

$$\langle f, g \rangle_{\mathbb{L}^2(\mathbb{R}^2)} = \left\langle \hat{f}, \hat{g} \right\rangle_{\mathbb{L}^2(\mathbb{R}^2)}, \quad (4.4)$$

où $\langle \cdot, \cdot \rangle_{\mathbb{L}^2(\mathbb{R}^2)}$ est le produit scalaire standard de l'espace $\mathbb{L}^2(\mathbb{R}^2)$. Comme la base de Fourier diagonalise les translations, et que les similitudes conservent le rapport des distances, la transformée de Fourier de la fonction $g_{a,\theta,b} = g \circ \phi_{a,\theta,b}$ ($b \in \mathbb{R}^2$, $a \in \mathbb{R}_+^*$ et $\theta \in \mathbb{R}$) est,

$$\hat{g}_{a,\theta,b}(\omega) = a^2 e^{i\omega \cdot b} \hat{g}(aA_\theta \omega), \quad \omega \in \mathbb{R}^2. \quad (4.5)$$

En particulier pour $a = 1$ et $\theta = 0$, nous avons que,

$$\hat{g}_{1,0,b}(\omega) = e^{i\omega \cdot b} \hat{g}(\omega), \quad \omega \in \mathbb{R}^2. \quad (4.6)$$

Ces deux dernières propriétés sont utilisées par la suite pour définir les fonctions de contraste.

4.2.2 Transformée de Fourier-Mellin

La transformée de Fourier Mellin est une transformée de type Fourier définie sur le plan polaire $\mathcal{G} = \mathbb{R}_+ \times \mathbb{T}$. Elle est utilisée en analyse harmonique, car elle traduit plus simplement des transformations géométriques telles que la rotation ou l'homothétie en des translations. Plusieurs auteurs ont développé des transformées de Fourier pour étudier des groupes de transformations spécifiques dans le contexte de l'analyse d'image 2D ou 3D. Parmi les nombreuses références sur ce sujet, on peut citer [Derrode et Ghorbel \(2004\)](#), [Ghorbel \(1994\)](#), [Lenz \(1990\)](#), [Gauthier et al. \(1991\)](#) et [Segman et al. \(1992\)](#).

Tout au long de ce chapitre, nous notons par $\mathbb{L}^p(\mathcal{G})$ l'espace des fonctions p -intégrables sur l'espace mesuré $(\mathcal{G}, \frac{dr}{r} \frac{d\theta}{2\pi})$. On muni cet espace de la norme suivante :

$$\|f\|_{L^p(\mathcal{G})} = \left(\int_0^{+\infty} \int_0^{2\pi} |f(r, \theta)|^p \frac{d\theta}{2\pi} \frac{dr}{r} \right)^{1/p}.$$

Soit $(k, v) \in \mathbb{Z} \times \mathbb{R}$, la transformée de Fourier-Mellin (FMT) d'une fonction $g \in \mathbb{L}^1(\mathcal{G})$ au point (k, v) est,

$$\tilde{\mathcal{M}}_g(k, v) = \int_0^{+\infty} \int_0^{2\pi} g(r, \theta) r^{-iv} e^{-ik\theta} \frac{d\theta}{2\pi} \frac{dr}{r}. \quad (4.7)$$

Comme l'a expliqué [Derrode et Ghorbel \(2004\)](#), la FMT peut être difficilement calculable dans la pratique puisque les images réelles n'appartiennent pas généralement à l'espace $\mathbb{L}^2(\mathcal{G})$. En effet, il suffit remarquer que la FMT existe pour toute fonction $g(r, \theta)$ qui admettent un équivalent de r^σ au voisinage de $r = 0$ pour une constante $\sigma > 0$. Ainsi, une image à niveau de gris ne peut pas remplir ce type de condition, puisqu'à proximité de $r = 0$, la fonction g est généralement non nulle. Dans la plupart des cas, une image g n'est donc pas intégrable pour la mesure $\frac{r}{dr} \frac{d\theta}{2\pi}$, et l'intégrale (4.7) diverge. Une solution, la plus souvent adoptée dans la pratique, est d'annuler l'image sur un petit disque au voisinage de l'origine. Or, pour notre modèle, le choix du rayon est problématique puisque l'on veut comparer la même image à des échelles différentes : une information non négligeable peut être perdue si on choisit un disque de rayon constant.

Afin de contourner cette difficulté, [Derrode et Ghorbel \(2004\)](#) et [Ghorbel \(1994\)](#) ont proposé de remplacer la fonction g par la fonction $g_\sigma(r, \theta) = r^\sigma g(r, \theta)$ pour une constante σ positive dont le choix est discuté plus tard. Cette pondération par un facteur σ définit la transformée de Fourier Mellin Analytique (AFMT) pour une fonction g telle que $g_\sigma \in \mathbb{L}^1(\mathcal{G})$:

$$\mathcal{M}_g(k, v) = \int_0^{+\infty} \int_0^{2\pi} g(r, \theta) r^{\sigma-iv} e^{-ik\theta} \frac{d\theta}{2\pi} \frac{dr}{r}, \quad \forall (k, v) \in \mathbb{Z} \times \mathbb{R}. \quad (4.8)$$

Si l'AFMT de g est intégrable,

$$\int_{\mathbb{R}} \sum_{k \in \mathbb{Z}} |\mathcal{M}_g(k, v)| dv < \infty,$$

alors l'AFMT est inversible, c'est-à-dire :

$$r^\sigma g(r, \theta) = \int_{-\infty}^{+\infty} \sum_{k \in \mathbb{Z}} \mathcal{M}_g(k, v) r^{iv} e^{ik\theta} dv, \quad \forall (r, \theta) \in \mathcal{G}. \quad (4.9)$$

D'autre part, si $g \in \mathbb{L}^1(\mathcal{G}) \cap \mathbb{L}^2(\mathcal{G})$, alors l'équation de Parseval est vérifiée :

$$\|g_\sigma\|_{L^2(\mathcal{G})}^2 = \int_{-\infty}^{+\infty} \sum_{k \in \mathbb{Z}} |\mathcal{M}_g(k, v)|^2 dv. \quad (4.10)$$

Les résultats (4.9) et (4.10) sont des résultats généraux pour des transformées de Fourier définies sur les groupes abéliens localement compact (voir Rudin (1990) pour plus de détails).

Maintenant, observons comment l'AFMT agit sur les rotations et sur les homothéties du plan. Soient $a \in \mathbb{R}_+^*$ et $\theta \in \mathbb{R}$. Si l'AFMT d'une fonction g est bien définie, alors on a :

$$\mathcal{M}_{g_{a,\theta,(0,0)}}(k, v) = a^{\sigma-iv} e^{ik\theta} \mathcal{M}_g(k, v), \quad \forall (k, v) \in \mathbb{Z} \times \mathbb{R}, \quad (4.11)$$

où $g_{a,\theta,(0,0)}$ la fonction suivante $g \circ \phi_{a,\theta,(0,0)}$. Par conséquent, alors que la transformée de Fourier standard convertit l'opérateur de translation par un paramètre de phase sans le domaine de Fourier, l'équation (4.11) prouve que l'AFMT convertit l'homothétie et la rotation en multiplication par un scalaire complexe dans le domaine de Fourier Mellin. L'AFMT est ainsi un outil adapté pour l'analyse d'images traitées dans ce chapitre.

4.3 Estimation des paramètres

À partir des résultats de Gamboa et al. (2005) et de ceux étudiés dans le chapitre 2, on aimerait proposer un processus de contraste qui estime la fonction de contraste suivante,

$$Q(\alpha) = \frac{1}{J} \sum_{j=1}^J \|f_j^* \circ \phi_{\frac{1}{a_j}, \tilde{b}_j, -\theta_j} - \frac{1}{J} \sum_{j'=1}^J f_{j'}^* \circ \phi_{\frac{1}{a_{j'}}, \tilde{b}_{j'}, -\theta_{j'}}\|_{L^2(\mathbb{R}^2)}^2, \quad (4.12)$$

où $\tilde{b}_j = -\frac{1}{a_j} A_{\theta_j} b_j$. Ce critère est étroitement lié à l'analyse Procruste (cette méthode est la représentation simultanée de différents spécimens par superposition) qui est classiquement employée pour l'analyse des formes (voir par exemple Dryden et Mardia (1998)) et qui a été proposé par Ramsay et Li (1998) pour l'alignement d'un ensemble de courbes sur une courbe de référence. Ici, l'image de référence est définie comme la moyenne des images synchronisées par les transformations $\phi_{a_j, \theta_j, b_j}$. Évidemment ce critère a un unique minimum en,

$$\alpha^* = (a_1^*, \dots, a_J^*, \theta_1^*, \dots, \theta_J^*, b_1^*, \dots, b_J^*),$$

tel que $M(\alpha^*) = 0$. Dans la pratique, ce sont des images bruitées qui sont observées, le processus de contraste que l'on pourrait considérer serait alors défini à partir de la fonction Q en remplaçant les fonctions f_j^* par les données Y_j . Cependant, d'un point de vue numérique, évaluer $Q(\alpha)$ nécessite le calcul des déformations des images f_j^* par les transformations $\phi_{a_j, \theta_j, b_j}$: ceci exige une interpolation des images. D'ailleurs une minimisation de $Q(\alpha)$ n'est certainement pas évidente, puisque c'est une fonction non linéaire de α . Un algorithme de type gradient exigerait par exemple l'évaluation des dérivées des images qui peuvent s'avérer assez difficile compte tenu de la présence du bruit.

La situation est beaucoup plus simple si nous supposons que tous les b_j^* sont nuls, c'est-à-dire que les images diffèrent par une rotation et une homothétie. Dans ce cas-ci, grâce à la propriété de l'AFMT (4.11) sur les homothéties et les rotations et grâce à la propriété de Parseval (4.10), nous pouvons récrire Q (en remplaçant l'espace $\mathbb{L}^2(\mathbb{R}^2)$ par $\mathbb{L}^2(\mathcal{G})$) sous la forme suivante,

$$Q_1(\alpha_1) = \frac{1}{J} \sum_{j=1}^J \int_{-\infty}^{+\infty} \sum_{k \in \mathbb{Z}} \left| a_j^{-\sigma+iv} e^{-ik\theta_j} \mathcal{M}_{f_j^*}(k, v) - \frac{1}{J} \sum_{j'=1}^J a_{j'}^{-\sigma+iv} e^{-ik\theta_{j'}} \mathcal{M}_{f_{j'}^*}(k, v) \right|^2 dv, \quad (4.13)$$

avec $\alpha_1 = (a_1, \dots, a_J, \theta_1, \dots, \theta_J)$.

De même, le critère Q se simplifie si nous supposons que tous les paramètres a_j^* sont égaux à 1 et que tous les paramètres θ_j^* sont nuls, c'est-à-dire que les images sont identiques à une translation près. En raison de la propriété (4.6) entre la transformée de Fourier et l'opérateur de translation, et de la propriété de Parseval (4.4), le critère Q peut se récrire sous la forme suivante,

$$Q_2(\alpha_2) = \frac{1}{J} \sum_{j=1}^J \int_{\mathbb{R}^2} \left| e^{i\omega \cdot b_j} \hat{f}_j^*(\omega) - \frac{1}{J} \sum_{j'=1}^J e^{i\omega \cdot b_{j'}} \hat{f}_{j'}^*(\omega) \right|^2 d\omega, \quad (4.14)$$

avec $\alpha_2 = (b_1, \dots, b_J)$.

Les critères Q_1 et Q_2 consistent à calculer l'AFMT et les transformées de Fourier des images f_j^* respectivement, puis de les normaliser afin d'aligner les images. Pour le critère Q_1 , on modifie les phases et les amplitudes des transformées de Fourier Mellin, tandis que seulement les phases des transformées de Fourier sont modifiées pour le critère Q_2 . Par conséquent, l'évaluation de ces deux critères est plus facile que le critère Q . De plus, à notre connaissance, aucune transformation ne possède simultanément une propriété du type (4.6) et (4.11) pour les translations, les rotations et les homothéties. Par conséquent, nous développons un procédé en deux étapes fondé sur deux processus de contraste similaires à Q_1 et Q_2 , et qui utilisent, eux aussi, la propriété

(4.11) de l'AFMT pour estimer les angles de rotation et les rapports d'homothétie, et la propriété (4.5) de la transformée de Fourier pour estimer les vecteurs de translation.

4.3.1 Un critère pour l'estimation simultanée des échelles et des angles de rotation

Utilisation combinée de la transformée de Fourier et de l'AFMT

Transformée de Fourier du modèle. Étant donné les hypothèses sur les paramètres a_j^* , θ_j^* et b_j^* , le support des fonctions f_j^* est inclus dans $\mathcal{D} = [-d/2, d/2]^2$. Ainsi on a,

$$\hat{f}_j(\omega) = \int_{\mathbb{R}^2} f_j^*(x) e^{-i\omega \cdot x} \frac{dx}{2\pi} = \int_{\mathcal{D}} f_j^*(x) e^{-i\omega \cdot x} \frac{dx}{2\pi}.$$

La transformée de Fourier du modèle est alors définie comme,

$$\hat{Y}_j(\omega) = \int_{\mathcal{D}} e^{-i\omega \cdot x} \frac{dY_j(x)}{2\pi} = \hat{f}_j(\omega) + \epsilon \hat{W}_j(\omega), \quad \omega \in \mathbb{R}^2, \quad (4.15)$$

avec $\hat{W}_j(\omega) = \int_{\mathcal{D}} e^{-i\omega \cdot x} \frac{dW_j(x)}{2\pi}$. Pour tout $\omega \in \mathbb{R}^2$, $\hat{W}_j(\omega)$ est un processus gaussien complexe de moyenne nulle, de variance,

$$\mathbb{E} \left| \hat{W}_j(\omega) \right|^2 = \int_{\mathcal{D}} dx = \frac{d^2}{(2\pi)^2},$$

et de covariance donnée par,

$$\mathbb{E} \left(\hat{W}_j(\omega) \overline{\hat{W}_j(\omega')} \right) = \int_{\mathcal{D}} e^{-i(\omega - \omega') \cdot x} \frac{d^2}{(2\pi)^2}, \quad \omega \in \mathbb{R}^2, \quad \omega' \in \mathbb{R}^2.$$

D'après la propriété (4.5), considérons maintenant le module carré de la transformée de Fourier afin d'éliminer le paramètre de translation. L'équation (4.15) devient,

$$|\hat{Y}_j|^2(\omega) = |\hat{f}_j|^2(\omega) + \epsilon w_j(\omega), \quad \omega \in \mathbb{R}^2, \quad (4.16)$$

avec pour tout $\omega \in \mathbb{R}^2$,

$$\begin{aligned} |\hat{f}_j|^2(\omega) &= |a_j^*|^4 |\hat{f}|^2(a_j^* A_{\theta_j^*} \omega), \\ w_j(\omega) &= 2\Re \left[\hat{f}_j(\omega) \overline{\hat{W}_j(\omega)} \right] + |\hat{W}_j|^2(\omega), \end{aligned}$$

où $\Re(c)$ désigne la partie réelle d'un nombre complexe c .

AFMT et critère limite. Afin de simplifier les notations, pour tout $(r, \theta) \in \mathcal{G}$, on associe le vecteur de ses coordonnées cartésiennes $\omega = (r \cos \theta, r \sin \theta) \in \mathbb{R}^2$ et vice versa. Remarquons que les fonctions $|\hat{f}_j|^2$ diffèrent des unes des autres par un coefficient d'amplitude a_j^{*4} , d'une homothétie de rapport a_j^* et d'une rotation d'angle θ_j^* . À condition que les fonctions $r^\sigma |\hat{f}_j|^2(r, \theta)$ appartiennent à l'espace $\mathbb{L}^2(\mathcal{G})$, on propose le critère limite suivant pour $\alpha_1 = (a_1, \dots, a_J, \theta_1, \dots, \theta_J) \in \mathcal{A}_1$,

$$M_1(\alpha_1) = \frac{1}{J} \sum_{j=1}^J \int_0^{+\infty} \int_0^{2\pi} \left| a_j^{-4} r^\sigma |\hat{f}_j|^2\left(\frac{r}{a_j}, \theta - \theta_j\right) - \frac{1}{J} \sum_{j'=1}^J a_{j'}^{-4} r^\sigma |\hat{f}_{j'}|^2\left(\frac{r}{a_{j'}}, \theta - \theta_{j'}\right) \right|^2 \frac{d\theta dr}{2\pi r},$$

avec,

$$\mathcal{A}_1 = \{(a_1, \dots, a_J, \theta_1, \dots, \theta_J) \in [a_{\min}, a_{\max}]^J \times [0, 2\pi]^J, \text{ tel que } a_1 = 1, \theta_1 = 0\}.$$

Bien évidemment, le critère M_1 admet un minimum en,

$$\alpha_1^* = (a_1^*, \dots, a_J^*, \theta_1^*, \dots, \theta_J^*)$$

tel que $M_1(\alpha_1^*) = 0$. La linéarité de l'AFMT et l'équation (4.11) (sous condition que l'AFMT soit bien définie pour les fonctions $r^\sigma |\hat{f}_j|(r, \theta)$) impliquent l'équation suivante,

$$\mathcal{M}_{|\hat{f}_j|^2}(k, v) = a_j^{*4-\sigma+iv} \mathcal{M}_{|\hat{f}_j|^2}(k, v) e^{ik\theta_j^*}, \quad v \in \mathbb{R}, \quad k \in \mathbb{Z}. \quad (4.17)$$

Ainsi par l'identité de Parseval pour la transformée de Fourier Mellin (4.10), le critère M_1 peut se récrire comme,

$$M_1(\alpha_1) = \frac{1}{J} \sum_{j=1}^J \int_{\mathbb{R}} \sum_{k \in \mathbb{Z}} \left| c_j(k, v) - \frac{1}{J} \sum_{j'=1}^J c_{j'}(k, v) \right|^2 dv, \quad (4.18)$$

avec

$$c_j(k, v) = a_j^{-4+\sigma-iv} \mathcal{M}_{|\hat{f}_j|^2}(k, v) e^{-ik\theta_j}. \quad (4.19)$$

Estimation des paramètres d'échelles et de rotation

Définition du processus de contraste. Dans la pratique, nous observons des versions bruitées des images f_j^* , nous proposons une méthode de M -estimation consistante pour le paramètre α_1^* . En raison du bruit, ce ne sont pas les transformée de Fourier Mellin des fonctions $r^\sigma |\hat{f}_j|(r, \theta)$ qui sont directement observées, mais des version bruitées,

$$\mathcal{M}_{|\hat{Y}_j|^2}^\epsilon(k, v) = \int_0^{\delta_\epsilon} \int_0^{2\pi} |\hat{Y}_j|^2(r, \theta) r^{\sigma-iv} e^{-ik\theta} \frac{d\theta dr}{2\pi r} \quad (4.20)$$

$$= \mathcal{M}_{|\hat{f}_j|^2}^\epsilon(k, v) + \epsilon \mathcal{M}_{w_j}^\epsilon(k, v), \quad (4.21)$$

où,

$$\begin{aligned}\mathcal{M}_{|\hat{f}_j|^2}^\epsilon(k, v) &= \int_0^{\delta_\epsilon} \int_0^{2\pi} |\hat{f}_j|^2(r, \theta) r^{\sigma-iv} e^{-ik\theta} \frac{d\theta}{2\pi} \frac{dr}{r} \\ \mathcal{M}_{w_j}^\epsilon(k, v) &= \int_0^{\delta_\epsilon} \int_0^{2\pi} w_j(r, \theta) r^{\sigma-iv} e^{-ik\theta} \frac{d\theta}{2\pi} \frac{dr}{r},\end{aligned}$$

et δ_ϵ est un réel positif de lissage dont la définition est précisée plus tard. Les coefficients c_j (4.19) sont estimés alors par,

$$d_j^\epsilon(k, v) = a_j^{-4+\sigma-iv} \mathcal{M}_{|\hat{Y}_j|^2}^\epsilon(k, v) e^{-ik\theta_j}.$$

Le processus de contraste proposé pour l'estimation de α_1^* est alors,

$$M_{1,\epsilon}(\alpha_1) = \frac{1}{J} \sum_{j=1}^J \int_{|v| \leq v_\epsilon} \sum_{|k| \leq k_\epsilon} \left| d_j^\epsilon(k, v) - \frac{1}{J} \sum_{j'=1}^J d_{j'}^\epsilon(k, v) \right|^2 dv, \quad (4.22)$$

où les réels k_ϵ et v_ϵ définissent des fenêtres sur le domaine de Fourier Mellin garantissant la convergence de $M_{1,\epsilon}$ vers M_1 .

Évidemment, la forme commune des images f ne doit pas être invariante par rotation, autrement on ne pourrait garantir l'unicité de la fonction de contraste M_1 . Par ailleurs, pour toute fonction réelle g , le module au carré de la transformée de Fourier est invariant par la rotation d'angle π :

$$|\hat{g}|^2(r, \theta) = |\hat{g}|^2(r, \theta + \pi), \quad \forall (r, \theta) \in \mathcal{G}.$$

Ainsi le critère est inchangé si on remplace le paramètre de rotation θ_j par $\theta_j + \pi$. Par conséquent, on minimise le processus de contraste non plus sur l'espace des paramètres \mathcal{A}_1 , mais sur,

$$\tilde{\mathcal{A}}_1 = \{(a_1, \dots, a_J, \theta_1, \dots, \theta_J) \in [a_{\min}, a_{\max}]^J \times [0, \pi]^J, \text{ tel que } a_1 = 1, \theta_1 = 0\},$$

afin de garantir l'unicité du minimum pour M_1 : nous estimons l'angle de rotation à π près et non à 2π près. Par ailleurs, si la fonction $|\hat{g}|^2$ est invariante par la rotation d'angle π , alors pour tout $k \in \mathbb{Z}$, les fonctions $v \rightarrow \mathcal{M}_{|\hat{g}|^2}(2k+1, v)$ sont identiquement nulles. L'invariance par rotation telle que nous l'avons comprise ici peut ainsi se traduire sur l'AFMT (c.f. la définition 4.1) du module au carré de \hat{f} .

Définition 4.1. Soit f une fonction de $\mathbb{L}^1(\mathbb{R}^2)$, telle que la fonction $r^\sigma |\hat{f}|^2(r, \theta)$ appartienne à l'espace $\mathbb{L}^1(\mathcal{G})$. La fonction f est dite non invariante par rotation s'il existe deux entiers premiers entre eux $(k, k') \in \mathbb{Z}^2$ tel que les fonctions $v \mapsto \mathcal{M}_{|\hat{f}|^2}(2k, v)$ et $v \mapsto \mathcal{M}_{|\hat{f}|^2}(2k', v)$ ne soient pas identiquement nulles.

Consistance des estimateurs. Si l'image f est assez régulière et n'est pas invariante par rotation, le théorème 4.1 assure la consistance de l'estimateur défini ci-dessous :

$$\hat{\alpha}_{1,\epsilon} = \arg \min_{\alpha_1 \in \hat{\mathcal{A}}_1} M_{1,\epsilon}(\alpha_1).$$

Avant d'énoncer le théorème, l'application de notre méthode d'estimation oblige à choisir trois paramètres $k_\epsilon \in \mathbb{N}^*$, et $(v_\epsilon, \delta_\epsilon) \in (\mathbb{R}_+^*)^2$. δ_ϵ est un filtre dans le domaine de Fourier utilisé pour calculer la transformée de Fourier Mellin des données. Les paramètres (k_ϵ, v_ϵ) sont deux filtres dans le domaine de Fourier Mellin intervenant dans la définition du processus de contraste $M_{1,\epsilon}$. Afin de garantir la convergence de la méthode, on doit supposer que ϵ converge vers zéro, et que ces derniers paramètres ont un comportement asymptotique spécifique.

Théorème 4.1. *Supposons que la forme commune f ne soit pas invariante par rotation. Soit $s > 0$, supposons que f vérifie les propriétés de régularité suivantes,*

$$\int_{v \in \mathbb{R}} \sum_{k \in \mathbb{Z}} |\mathcal{M}_{|\hat{f}|^2}(k, v)|^2 dv < \infty \quad (4.23)$$

$$\int_{\omega \in \mathbb{R}^2} |\hat{f}|^2(\omega) |\omega|^{2s} d\omega < \infty \quad (4.24)$$

$$\sigma < 2s + 2 \quad (4.25)$$

Supposons également que k_ϵ , v_ϵ , δ_ϵ et ω_ϵ sont tels que $\epsilon \rightarrow 0$

$$k_\epsilon, v_\epsilon, \delta_\epsilon \rightarrow +\infty \quad (4.26)$$

$$\delta_\epsilon^{-4-4s+2\sigma} k_\epsilon v_\epsilon = \mathcal{O}(1) \quad (4.27)$$

$$\epsilon^2 \delta_\epsilon^\sigma k_\epsilon v_\epsilon = o(1). \quad (4.28)$$

Alors $\hat{\alpha}_{1,\epsilon}$ converge en probabilité vers α^ .*

Quelques remarques au sujet des conditions du théorème 4.1 peuvent être faites.

- i) Puisque f est à support compact, on a bien que $f \in \mathbb{L}^1(\mathbb{R}^2) \cap \mathbb{L}^2(\mathbb{R}^2)$, il existe donc $\sigma \leq 2$ tel que la fonction $r^\sigma |\hat{f}|^2(r, \theta)$ appartienne à $\mathbb{L}^1(\mathcal{G})$. L'AFMT de $|\hat{f}|^2$ est donc bien définie.
- ii) L'hypothèse (4.23) signifie que la fonction $r^\sigma |\hat{f}|^2(r, \theta)$ appartient à $\mathbb{L}^2(\mathcal{G})$ d'après l'égalité de Parseval (4.10),

$$\int_{v \in \mathbb{R}} \sum_{k \in \mathbb{Z}} |\mathcal{M}_{|\hat{f}|^2}(k, v)|^2 dv = \int_{\omega \in \mathbb{R}^2} |\hat{f}(\omega)|^4 |\omega|^{2\sigma-2} \frac{d\omega}{2\pi}.$$

D'autre part, puisque f est de carré intégrable, l'auto-convolution de f ,

$$x \in \mathbb{R}^2 \rightarrow \int_{\omega \in \mathbb{R}^2} f(y)f(y-x)dy. \quad (4.29)$$

est intégrable sur \mathbb{R}^2 , et sa transformée de Fourier est $|\hat{f}(\omega)|^2$. Ainsi, quand $\sigma \geq 1$, l'hypothèse (4.23) signifie que la fonction (4.29) appartient dans un espace de Sobolev d'ordre $\sigma - 1$. Quand $\sigma < 1$, l'hypothèse (4.23) est une conséquence de la condition (4.24).

- iii) L'hypothèse (4.24) est une hypothèse classique sur f : elle appartient à un espace de Sobolev d'ordre $s > 0$.
- iv) L'hypothèse (4.25) implique que σ ne doit pas être plus grand que la régularité de la fonction f mesurée par la paramètre s . Cette hypothèse est utilisée dans la preuve du théorème 4.1 afin de garantir la convergence de certaines intégrales.
- v) Finalement, les conditions (4.27) et (4.28) contrôlent la vitesse de convergence vers l'infini des paramètres de fenêtre k_ϵ , v_ϵ et δ_ϵ , quand $\epsilon \rightarrow 0$, afin d'assurer l'uniforme convergence en probabilité de $M_{1,\epsilon}$ vers M_1 . Par exemple, si $k_\epsilon = \epsilon^{-1+\sigma/(2+2s)}$, $v_\epsilon = \epsilon^{-1+\sigma/(2+2s)}$ et $\delta_\epsilon = \epsilon^{-1/(2+2s)}$, les hypothèses (4.27) et (4.28) sont vérifiées.

Remarque 4.1. *On peut mettre en parallèle les hypothèses du théorème 2.3 du chapitre 2 et celles du théorème 4.1,*

- *la condition (4.24) est l'équivalent de l'hypothèse sur la vitesse de la convergence de la série de Fourier ;*
- *dans le chapitre 2, on avait $\epsilon = 1/n$, ainsi l'hypothèse (4.28) correspond bien à l'hypothèse $m_n^2/n = o(1)$, sauf que la troncature δ_ϵ sur le domaine de Fourier intervient. Ceci vient du fait que le domaine d'intégration ($\mathcal{D} = \mathbb{T}^2$) au chapitre 2 est compact, alors que le domaine considéré ici est le domaine de Fourier \mathbb{R}^2 ;*
- *les conditions (4.25) et (4.27) mettent en relation la régularité de la forme commune et des fenêtres k_ϵ et v_ϵ sur le domaine de Fourier Mellin. Ces hypothèses n'apparaissent pas explicitement dans le théorème 2.3, mais elles interviennent dans le choix explicite de la régularité de la forme commune et la propriété d'orthogonalité de la base de Fourier discrète (voir le lemme 2.5 ainsi que le commentaire qui le précède).*

D'ailleurs, avec des conditions plus fortes, le théorème suivant affirme que notre méthode fournit des estimateurs ϵ^{-1} -consistants. On peut établir que la loi limite des fluctuations : c'est une variable gaussienne centrée. La matrice de covariance étant trop compliquée, elle n'est pas explicitement calculée.

Théorème 4.2. Notons par g la fonction $r^\sigma |\hat{f}|^2(r, \theta)$. Supposons que g soit différentiable telle que ses dérivées partielles $\partial_\theta g$ et $\partial_r g$ relativement aux coordonnées polaire r et θ appartiennent à l'espace $\mathbb{L}^2(\mathcal{G})$, et que,

$$\int_{v \in \mathbb{R}} \sum_{k \in \mathbb{Z}} (|k| + |v|) |\mathcal{M}_{|\hat{f}|^2}(k, v)| dv < \infty \quad (4.30)$$

$$\int_{\omega \in \mathbb{R}^2} |\hat{f}|^2(\omega) |\omega|^{2s} d\omega < \infty \quad (4.31)$$

$$\sigma \neq 4 \text{ et } 3\sigma/2 < 2s + 2, \quad (4.32)$$

et que f , $\partial_\theta g$ ne soient pas identiquement nulles. De plus, nous supposons que,

$$(\sigma - 5)(\sigma - 4) \|g\|_{L^2(\mathcal{G})}^2 + \|\partial_r g\|_{L^2(\mathcal{G})}^2 \neq 0, \quad \text{et} \quad (\sigma - 4)^2 \frac{1}{J} - (\sigma - 4) > 0. \quad (4.33)$$

Supposons que k_ϵ , v_ϵ et δ_ϵ sont tels que quand $\epsilon \rightarrow 0$,

$$k_\epsilon, v_\epsilon, \delta_\epsilon, \omega_\epsilon \rightarrow +\infty \quad (4.34)$$

$$\delta_\epsilon^{-4-4s+2\sigma} k_\epsilon v_\epsilon (k_\epsilon + v_\epsilon) \epsilon^{-1} = \mathcal{O}(1) \quad \text{et} \quad \delta_\epsilon^{-2-2s+\sigma} \epsilon^{-1} = \mathcal{O}(1) \quad (4.35)$$

$$\epsilon k_\epsilon v_\epsilon (k_\epsilon + v_\epsilon) \delta_\epsilon^\sigma = o(1), \quad (4.36)$$

$$k_\epsilon v_\epsilon (k_\epsilon + v_\epsilon) \delta_\epsilon^{-2s-2+3\sigma/2} = \mathcal{O}(1). \quad (4.37)$$

Alors $\epsilon^{-1}(\hat{\alpha}_{1,\epsilon} - \alpha_1^*)$ converge en loi vers un vecteur gaussien centré.

De même que pour le théorème précédent, quelques commentaires sur les hypothèses de théorème (4.2) peuvent être faites.

- i) L'hypothèse (4.30) signifie que les transformées de Fourier Mellin de $\partial_\theta g$ et $\partial_r g$ sont intégrables dans le domaine de Fourier Mellin. Ainsi on peut utiliser la formule inverse de la transformée de Fourier-Mellin (4.9) pour $\partial_r g$ et $\partial_\theta g$. Les fonctions $\partial_r g$ et $\partial_\theta g$ sont donc continues et uniformément bornées. Remarquons aussi que l'on ne suppose pas que la dérivée partielle en r , $\partial_r g$, soit non nulle. En effet, dans le modèle étudié (4.16), cette hypothèse n'est pas nécessaire puisque les paramètres a_j^* apparaissent en facteur de \hat{f}_j et dans la composante radiale.
- ii) La condition (4.32) assure que la fonction $r^\sigma |\hat{f}|^2(r, \theta)$ soit dans $\mathbb{L}^2(\mathcal{G})$.
- iii) La condition (4.33) permet que chaque coordonnée de α_1^* soit estimée à la vitesse ϵ^{-1} . Cette hypothèse est vérifiée par exemple pour $\sigma < 4$.
- iv) Enfin, les hypothèses (4.35)-(4.37) contrôlent la vitesse de convergence vers l'infini des paramètres k_ϵ , v_ϵ et δ_ϵ quand $\epsilon \rightarrow 0$, afin que les dérivées premières et secondes de $M_{1,\epsilon}$ convergent. Par exemple, si $\delta_\epsilon = \epsilon^{-1/(2+2s-\sigma/2)}$ et $k_\epsilon = \delta_\epsilon^\gamma$, $v_\epsilon = \delta_\epsilon^\gamma$ avec $2s + 2 - 3\sigma/2 > 3\gamma$, les hypothèses (4.35-4.37) sont vérifiées.

4.3.2 Un critère pour l'estimation simultanée des vecteurs de translation

Dans la section précédente, nous avons construit un estimateur ϵ^{-1} -consistant $\hat{\alpha}^{1,\epsilon} = (\hat{a}_1, \dots, \hat{a}_J, \hat{\theta}_1, \dots, \hat{\theta}_J)$ de α_1^* . Pour $\alpha_2 = (b_1, \dots, b_J) \in \mathcal{A}_2$, ces estimateurs sont utilisés pour aligner les images à partir du modèle (4.15) de la façon suivante,

$$\begin{aligned} \hat{Z}_j(\omega) &= \hat{a}_j^{-2} \hat{Y}_j(\hat{a}_j^{-1} A_{-\hat{\theta}_j} \omega) e^{i\omega \tilde{\beta}_j} \\ &= \left(\frac{a_j^*}{\hat{a}_j} \right)^2 \hat{f} \left(\frac{a_j^*}{\hat{a}_j} A_{\theta_j^* - \hat{\theta}_j} \omega \right) e^{i\omega(\tilde{\beta}_j - \tilde{\beta}_j^*)} + \epsilon \hat{a}_j^{-2} \hat{W}_j(\hat{a}_j^{-1} A_{-\hat{\theta}_j} \omega) e^{i\omega \tilde{\beta}_j}, \end{aligned} \quad (4.38)$$

où les notations suivantes sont utilisées,

$$\begin{aligned} \beta_j &= a_j^{*-1} A_{\theta_j^*} b_j, & \beta_j^* &= a_j^{*-1} A_{\theta_j^*} b_j^* \\ \tilde{\beta}_j &= \hat{a}_j^{-1} A_{\hat{\theta}_j} b_j, & \tilde{\beta}_j^* &= \hat{a}_j^{-1} A_{\hat{\theta}_j} b_j^*. \end{aligned}$$

À partir de l'alignement (4.38), on définit le processus de contraste $M_{2,\epsilon}$:

$$M_{2,\epsilon}(\alpha_2) = \frac{1}{J} \sum_{j=1}^J \int_{|\omega| \leq \omega_\epsilon} \left| \hat{Z}_j(\omega) - \frac{1}{J} \sum_{j'=1}^J \hat{Z}_{j'}(\omega) \right|^2, \quad \alpha_2 \in \mathcal{A}_2, \quad (4.39)$$

où ω_ϵ est une fenêtre dans le domaine de Fourier. Le théorème 4.3 établit la consistance du M -estimateur défini comme le minimum de la fonction $M_{2,\epsilon}$:

$$\hat{\alpha}_2^\epsilon = \arg \min_{\alpha_2 \in \mathcal{A}_2} M_{2,\epsilon}(\alpha_2).$$

Théorème 4.3. *Supposons que les hypothèses du théorème 4.2 soient vérifiées et que la fonction \hat{f} soit 1-Lipschitz, c'est-à-dire qu'il existe $C > 0$ telle que pour tout $\omega, \omega' \in \mathbb{R}^2$,*

$$|\hat{f}(\omega) - \hat{f}(\omega')| \leq C \|\omega - \omega'\|.$$

Supposons que ω_ϵ soit tel que quand $\epsilon \rightarrow 0$,

$$\omega_\epsilon \rightarrow +\infty \quad (4.40)$$

$$\epsilon^2 \omega_\epsilon^3 = o(1). \quad (4.41)$$

Alors $\hat{\alpha}_{2,\epsilon}$ converge en probabilité vers α_2^ .*

Avec des hypothèses plus fortes, le théorème suivant établit la ϵ^{-1} -consistance de $\hat{\alpha}_{2,\epsilon}$. Une étude plus détaillée permettrait sans doute de déterminer plus explicitement la loi limite des fluctuations de l'estimateur des translations. Comme elle dépendrait de la loi limite des fluctuations $\hat{\alpha}_{1,\epsilon}$, elle serait néanmoins compliquée.

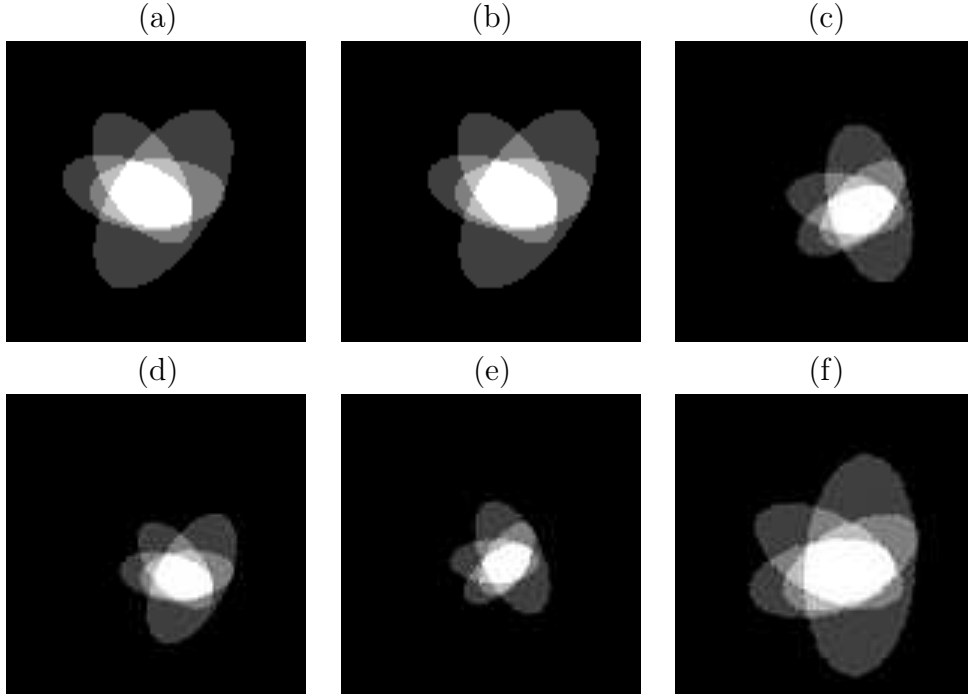


FIG. 4.2: Images simulées f_j^* : (a) $j = 1$, (b) $j = 2$, (c) $j = 3$, (d) $j = 4$, (e) $j = 5$, (f) $j = 6$.

Théorème 4.4. *Supposons que les hypothèses du théorème 4.2 soient vérifiées. On note respectivement par $\partial_1 f$ et $\partial_2 f$ les dérivées partielles de la fonction f par rapport à la première coordonnée cartésienne et la seconde coordonnée cartésienne. Supposons qu'il existe $\gamma > 0$, tel que \hat{f} soit 1-Lipchitz telle que,*

$$\partial_1 f \in \mathbb{L}^2(\mathcal{D}), \quad \partial_2 f \in \mathbb{L}^2(\mathcal{D}), \quad (4.42)$$

$$\int_{\omega \in \mathbb{R}^2} |\hat{f}|(\omega) |\omega|^2 d\omega < \infty, \quad (4.43)$$

et que $\partial_1 f$ et $\partial_2 f$ soient linéairement indépendantes dans l'espace vectorielle $\mathbb{L}^2(\mathcal{D})$. De plus supposons que ω_ϵ soit tel que $\epsilon \rightarrow 0$

$$\omega_\epsilon \rightarrow +\infty \quad (4.44)$$

$$\epsilon \omega_\epsilon^4 = o(1). \quad (4.45)$$

Alors $\epsilon^{-1}(\hat{\alpha}_{2,\epsilon} - \alpha_2^)$ est uniformément tendu.*

4.4 Simulations

Dans cette section, nous rapportons les résultats de quelques simulations pour étudier les exécutions numériques de notre procédé en deux étapes. Toutes les simulations ont été effectuées avec `Matlab`. Les données ont été générées à partir d'une image carrée simulée f de taille $N \times N$ ($N = 100$) que l'on a déformée. La figure 4.2 représente $J = 6$ versions déformées de cette image. Les paramètres des homothéties, des rotations et des translations sont reportées dans le tableau 4.1.

Image f_j^*	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$
a_j^*	1	1	0.8	0.7	0.6	1.1
b_j^*	(0,0)	(10,0)	(12,4)	(10,10)	(5,5)	(8,6)
θ_j^*	0	0	0.8	0.1	1	0.5

TAB. 4.1: Valeur des paramètres d'échelle, de rotation et de translation pour les $J = 6$ images

Pour simplifier, on note la valeur de l'image f_j^* au pixel (p_1, p_2) par $f_j^*(p_1, p_2)$ pour $p_1 = -(N-1)/2, \dots, (N-1)/2$ et $p_2 = -(N-1)/2, \dots, (N-1)/2$. Les images bruitées Y_{p_1, p_2}^j sont générées à partir du modèle de régression suivant,

$$Y_{p_1, p_2}^j = f_j^*(p_1, p_2) + \tau z_{p_1, p_2}^j,$$

où τ est un niveau de bruit inconnu, et z_{p_1, p_2}^j , $j = 1, \dots, J$, $p_1, p_2 = -(N-1)/2, \dots, (N-1)/2$, sont des variables gaussiennes indépendantes identiquement distribuées centrées et de variance égale à 1. Une quantité importante dans les simulations est la racine carrée du rapport signal-sur-bruit défini par

$$s2n = std(f)/\tau,$$

où $std(f)$ est l'écart type empirique de valeurs des pixels de l'image f . Nous présentons les résultats pour l'estimation des paramètres d'échelles, de rotation et de translation pour $s2n = 5, 3, 1$ (respectivement un niveau de bruit très faible, modéré et élevé). La figure 4.3 représente les images de la figure 4.2 bruitées avec $s2n = 1$.

Pour calculer la transformée de Fourier Mellin d'une image discrète g définie sur une grille cartésienne de taille $N \times N$, nous avons choisi de calculer d'abord sa transformée de Fourier \hat{g} sur une grille polaire de (r_{q_1}, θ_{q_2}) avec,

$$r_{q_1} = \frac{q_1}{N}, \quad \theta_{q_2} = 2\pi \frac{q_2 - 1}{N},$$

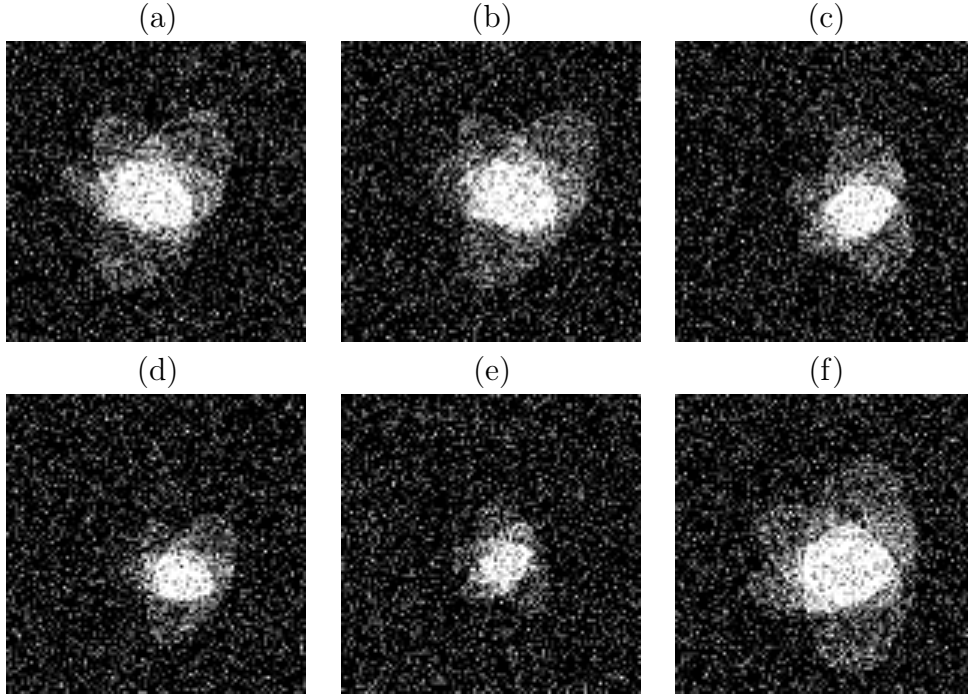


FIG. 4.3: Images bruitées avec $s2n = 1$: (a) $j = 1$, (b) $j = 2$, (c) $j = 3$, (d) $j = 4$, (e) $j = 5$, (f) $j = 6$.

et

$$\hat{g}(r_{q_1}, \theta_{q_2}) = \sum_{p_1=-(N-1)/2}^{(N-1)/2} \sum_{p_2=-(N-1)/2}^{(N-1)/2} e^{-i(p_1 r_{q_1} \cos(\theta_{q_2}) + p_2 r_{q_1} \sin(\theta_{q_2}))} g(p_1, p_2)$$

pour $q_1 = 1, \dots, N$ et $q_2 = 1, \dots, N$. Des implémentations plus rapides et plus précises ont été développées récemment pour calculer la transformée de Fourier polaire discrète (Averbuch et al. (2006)). Le but de nos simulations étant d'illustrer les bonnes propriétés de notre méthode en terme d'évaluation plutôt que de fournir un algorithme rapide, nous n'avons pas étudié cette dernière possibilité.

Pour tout $(k, v) \in \mathbb{Z} \times \mathbb{R}$, une approximation numérique de l'AFMT de $|\hat{g}|^2$ est obtenue de façon suivante,

$$\mathcal{M}_{|\hat{g}|^2}(k, v) = \sum_{q_1=1}^N \sum_{q_2=1}^N |\hat{g}|^2(r_{q_1}, \theta_{q_2}) r_{q_1}^{\sigma-1-iv} e^{-ik\theta_{q_2}}.$$

Pour la simplifier, nous avons choisi $\sigma = 1$, ceci afin de garantir les conditions (4.25) et (4.32) des théorèmes 4.1 et 4.2 pour tout $s > 0$. Il ne reste donc plus qu'à choisir les paramètres des différentes fenêtres k_ϵ , v_ϵ et δ_ϵ pour calculer les critères de $M_{1,\epsilon}$ défini par l'équation (4.22). Pour les simulations présentées, nous avons choisi $\delta_\epsilon = 1$. Ensuite,

pour les choix de $k_\epsilon = 5$ et de $v_\epsilon = 5$, nous avons constaté que les résultats étaient satisfaisants pour différentes valeurs du rapport signal-sur-bruit. Naturellement, il aurait été préférable d'avoir un critère de choix plus précis pour ces paramètres de fenêtre. Néanmoins ces petites valeurs semblent être un choix raisonnable.

Une fois ces paramètres choisis, et les AFMT des modules au carré des transformées de Fourier des images discrètes bruitées Y_j , $j = 1, \dots, J$ calculées, le critère $M_{1,\epsilon}$ peut être numériquement calculé pour tout vecteur $\alpha_1 \in \tilde{\mathcal{A}}_1$. Pour trouver le minimum de la fonction $M_{1,\epsilon}(\alpha_1)$, nous employons la méthode de minimisation de [Nelder et Mead \(1965\)](#) en utilisant la fonction `fminsearch` de `Matlab`. D'autres méthodes d'optimisation plus sophistiquées par exemple basées sur une descente de gradient auraient pu être considérées, mais nous avons constaté que cette approche simple donne de bons résultats.

Une fois les estimateurs \hat{a}_j et $\hat{\theta}_j$ pour les paramètres d'échelle et de rotation (avec les contraintes d'identifiabilité $\hat{a}_1 = 1$ et $\hat{\theta}_1 = 0$) évalués, les transformées de Fourier alignées $\hat{a}_j^{-2} \hat{Y}^j(\hat{a}_j^{-1} A_{-\hat{\theta}_j} \omega_{q_1, q_2})$ sont calculées (où $\omega_{q_1, q_2} = (r_{q_1} \cos(\theta_{q_2}), r_{q_1} \sin(\theta_{q_2}))$) pour $q_1, q_2 = 1, \dots, N$. Pour cela, on utilise des splines cubiques vérifiant des conditions d'annulation au bord du domaine polaire de Fourier. Pour tout paramètre $\alpha_2 \in \mathcal{A}_2$ de translation, le critère $M_{2,\epsilon}(\alpha_2)$ peut être numériquement évalué en approchant l'intégrale sur l'ensemble de fréquences tels que $|\omega| \leq \omega_\epsilon$ de (4.39) par une formule de quadrature (la somme de Riemann par exemple) sur les points ω_{q_1, q_2} pour $q_1 = 1, \dots, q_{\text{epsilon}}$ et $q_2 = 1, \dots, N$. Dans nos simulations, nous avons constaté que le choix $q_\epsilon = 3$ donne de bons résultats qui correspond au choix $\omega_{\text{epsilon}} = 3/N$. La méthode d'optimisation employée est la même méthode précédemment citée pour trouver le minimum de $M_{2,\epsilon}(\alpha_2)$.

Pour chaque valeur du rapport signal-sur-bruit $s2n$, nous avons simulé $M = 50$ jeux de données de J images bruitées. Pour chaque échantillon, les paramètres d'échelle, de rotation et de translation de chaque image ont été estimés. Le tableau 4.2 donne l'écart-type empirique et la moyenne empirique de ces estimateurs calculés à partir des estimateurs calculés pour les $M = 50$ jeux de données. Les paramètres de la première image étant fixé par les contraintes d'identifiabilité, seulement les résultats pour les images $j = 2, \dots, 6$ sont donnés

Les résultats présentés dans la table 4.2 montre que notre procédure estime plutôt bien les différents paramètres associées aux différentes images. Les erreurs standards sont inversement proportionnelles à $s2n$, et restent assez petites. Concernant les moyennes empiriques, elles sont assez proches des valeurs estimées. Cependant, il semble que certains estimateurs soient légèrement biaisés (par exemple $b_j^{1,*}$ pour $j = 3$ et $\hat{\theta}_j$ pour $j = 5$). Il serait intéressant d'étudier cette question : ce biais pourrait

TAB. 4.2: Moyennes empiriques et écarts-type (entre parenthèses) des estimateurs $\hat{a}_j, \hat{b}_j = (\hat{b}_j^1, \hat{b}_j^2), \hat{\theta}_j$ sur $M = 50$ jeux de données pour différent rapport signal-sur-bruit $s2n$. Les valeurs en gras représentent les valeurs à estimer des paramètres.

	$s2n$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$
a_j^*		1	0.8	0.7	0.6	1.1
\hat{a}_j	1	1.02 (0.00835)	0.819 (0.00945)	0.719 (0.0121)	0.618 (0.00811)	1.12 (0.0106)
	3	1.01 (0.00373)	0.814 (0.00329)	0.714 (0.0032)	0.613 (0.00347)	1.12 (0.0037)
	5	1.02 (0.00215)	0.815 (0.00164)	0.714 (0.0017)	0.612 (0.00185)	1.12 (0.0026)
$b_j^{1,*}$		10	12	10	5	8
\hat{b}_j^1	1	9.64 (1.92)	12.2 (1.05)	9.85 (1.24)	5.16 (1.36)	7.79 (1.69)
	3	9.69 (1.01)	12.3 (0.324)	9.98 (0.341)	5.34 (0.419)	8.29 (0.32)
	5	9.89 (0.49)	12.3 (0.215)	10 (0.159)	5.25 (0.298)	8.27 (0.19)
b_j^2		0	4	10	5	6
$\hat{b}_j^{2,*}$	1	0.107 (1.02)	5.58 (1.41)	10.6 (0.793)	6.36 (1.23)	7.46 (1.47)
	3	0.00603 (0.508)	5.75 (0.493)	10.2 (0.43)	6.53 (0.466)	7.53 (0.516)
	5	0.0217 (0.448)	5.68 (0.409)	10.2 (0.32)	6.52 (0.323)	7.58 (0.444)
θ_j^*		0	0.8	0.1	1	0.5
$\hat{\theta}_j$	1	0.197 (0.641)	0.7 (0.407)	0.281 (0.656)	0.834 (0.487)	0.554 (0.64)
	3	0.27 (0.935)	0.67 (0.324)	0.016 (0.273)	0.797 (0.398)	0.404 (0.284)
	5	0.0642 (0.501)	0.659 (0.295)	0.0652 (0.225)	0.837 (0.356)	0.428 (0.254)

provenir soit de la méthode numérique d'optimisation ou soit de la réutilisation des estimateurs des paramètres d'échelle et de rotation pour l'estimation des paramètres de translation.

4.4.1 Un exemple réel

On revient à l'exemple réel présenté au début de ce chapitre. Nous utilisons la même méthodologie décrite ci-dessus avec les mêmes paramètres pour les filtres k_ϵ , v_ϵ et δ_ϵ . Les trois images de la première ligne de la figure (4.1) sont alignées en choisissant comme image de référence (c'est-à-dire l'image n°1) la figure 4.1(a) : l'alignement est effectué par rapport à l'échelle, l'orientation et l'emplacement de ce papillon. De la même façon, on aligne les images de la deuxième ligne de la figure 4.1. L'image de référence (image n°1) est ici la figure 4.1(d). La figure 4.4 présente les images alignées en utilisant l'interpolation linéaire. Comme on peut l'observer les résultats sont très satisfaisants, puisque on distingue aucune différence entre les images après alignement.

4.5 Preuves

Dans les preuves, C désigne une constante arbitraire dont la valeur peut varier d'une ligne à une autre.

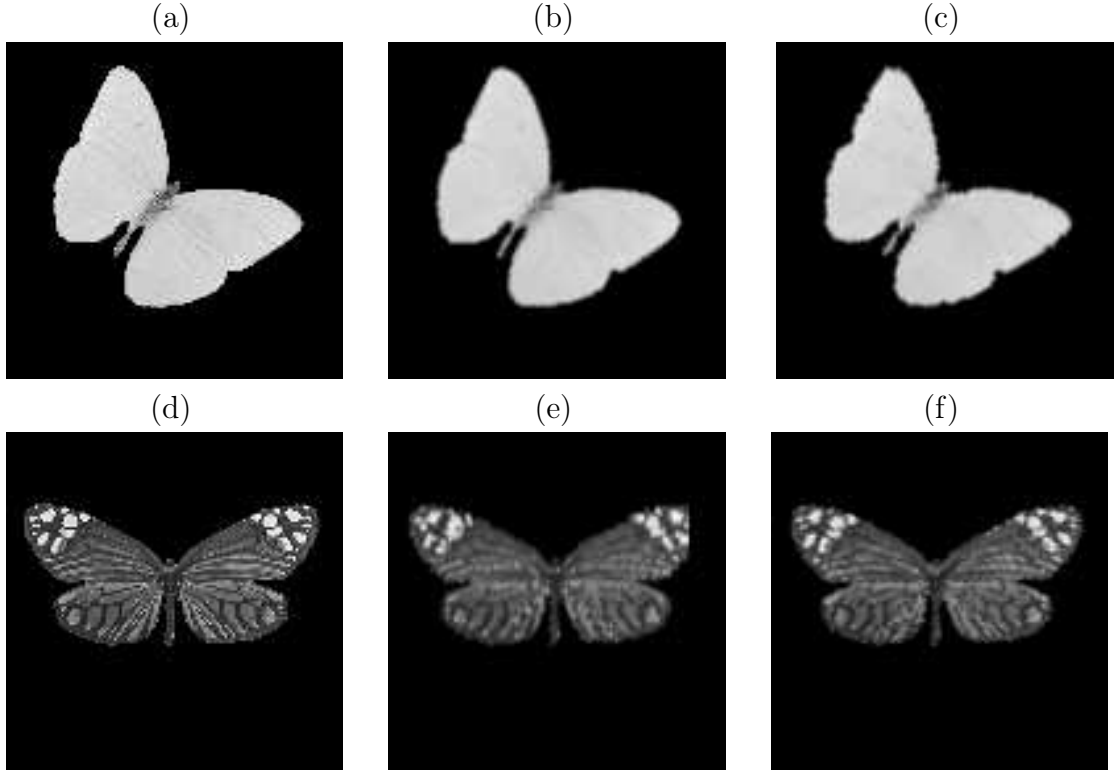


FIG. 4.4: Aligment des images de papillons.

Preuve du théorème 4.1

La preuve de ce théorème suit les lignes classiques de la convergence des M -estimateurs. On montre que $M_1(\alpha_1)$ admet un unique minimum en $\alpha_1 = \alpha_1^*$, et que $M_{1,\epsilon}$ converge uniformément (en α_1) en probabilité vers M_1 i.e.

$$\sup_{\alpha_1 \in \tilde{\mathcal{A}}_1} |M_{1,\epsilon}(\alpha_1) - M_1(\alpha_1)| = o_{\mathbb{P}}(1).$$

Ces deux conditions assurent que $\hat{\alpha}_{1,\epsilon}$ converge en probabilité vers α_1^* quand $\epsilon \rightarrow 0$ (voir le chapitre 1).

Unicité du minimum de $M(\alpha_1)$:

On rappelle que pour $\alpha_1 \in \tilde{\mathcal{A}}_1$,

$$M_1(\alpha_1) = \int_{\mathbb{R}} \sum_{k \in \mathbb{Z}} \left| \mathcal{M}_{|\hat{f}|^2}(k, v) \right|^2 \frac{1}{J} \sum_{j=1}^J \left| e^{ik(\theta_j^* - \theta_j)} \left(\frac{a_j}{a_j^*} \right)^{-4+\sigma-iv} - \frac{1}{J} \sum_{l=1}^J \left(\frac{a_l}{a_l^*} \right)^{-4+\sigma-iv} e^{ik(\theta_l^* - \theta_l)} \right|^2 dv.$$

Ainsi M_1 a un minimum en α_1^* car $M(\alpha_1^*) = 0$. Supposons qu'il existe $\alpha_1 \in \tilde{\mathcal{A}}_1$ tel que $M_1(\alpha_1) = 0$. Puisque \hat{f} est une fonction non nulle, l'ensemble des $(k, v) \in \mathbb{Z} \times \mathbb{R}$ tel que $\mathcal{M}_{|\hat{f}|^2}(k, v) \neq 0$ est non vide.

Soit $(k, v) \in \mathbb{Z} \times \mathbb{R}$ tel que $\mathcal{M}_{|\hat{f}|^2}(k, v) \neq 0$. Alors nous avons :

$$\frac{1}{J} \sum_{j=1}^J \left| \left(\frac{a_j}{a_j^*} \right)^{-4+\sigma-iv} e^{ik(\theta_j^*-\theta_j)} - \frac{1}{J} \sum_{l=1}^J \left(\frac{a_l}{a_l^*} \right)^{-4+\sigma-iv} e^{ik(\theta_l^*-\theta_l)} \right|^2 = 0.$$

Ceci signifie qu'il existe $\lambda_{k,v} \in \mathbb{C}$ tel que,

$$\left(\frac{a_j}{a_j^*} \right)^{-4+\sigma-iv} e^{ik(\theta_j^*-\theta_j)} = \lambda_{k,v}, \quad \forall j = 1 \dots J. \quad (4.46)$$

Or par les contraintes d'identifiabilité posées sur le modèle ($a_1 = a_1^*$ et $\theta_1 = \theta_1^* = 0$), on déduit que $\lambda_{k,v} = 1$. Si $\sigma \neq 4$, en considérant la module, l'égalité (4.46) devient :

$$\left(\frac{a_j}{a_j^*} \right)^{-4+\sigma} = 1, \quad \forall j = 1 \dots J,$$

et donc pour tout $j = 1 \dots J$, a_j est égal à a_j^* . Si $\sigma = 4$, puisque la fonction $v \in \mathbb{R} \rightarrow \mathcal{M}_{|\hat{f}|^2}(k, v)$ est continue et identiquement non nulle, il existe deux réels v_0 et v_1 tels que $v - v_0 \in \mathbb{Q}$, $v - v_1 \in \mathbb{R} \setminus \mathbb{Q}$ et $\mathcal{M}_{|\hat{f}|^2}(k, v_0) \neq 0$, $\mathcal{M}_{|\hat{f}|^2}(k, v_1) \neq 0$ (on utilise le fait que les espaces \mathbb{Q} et $\mathbb{R} \setminus \mathbb{Q}$ sont des sous-ensembles denses de \mathbb{R}). L'équation (4.46) est donc vérifiée pour v_0 et v_1 , et on a :

$$e^{-i(v-v_0)\log(a_j/a_j^*)} = 1 \quad \text{et} \quad e^{-i(v-v_1)\log(a_j/a_j^*)} = 1, \quad j = 1 \dots J.$$

Les ensembles \mathbb{Q} et $\mathbb{R} \setminus \mathbb{Q}$ étant par définition disjoints, on déduit que nécessairement pour tout $j = 1 \dots J$, $\log(a_j/a_j^*) = 0$, c'est-à-dire que a_j est égal à a_j^* .

Par ailleurs, la fonction f étant invariante par rotation, il existe deux entiers relatifs k et k' premiers entre eux tels que les fonctions $v \rightarrow \mathcal{M}_{|\hat{f}|^2}(2k, v)$ et $v \rightarrow \mathcal{M}_{|\hat{f}|^2}(2k', v)$ ne soient pas identiquement nulles. L'équation (4.46) implique alors que pour tout $j = 1, \dots, J$,

$$e^{i2k(\theta_j^*-\theta_j)} = 1 \quad \text{et} \quad e^{i2k'(\theta_j^*-\theta_j)} = 1.$$

Par l'absurde, supposons qu'il existe $j \geq 2$ tel que $\theta_j^* - \theta_j \neq 0$, alors l'équation précédente implique qu'il existe deux entiers p et p' tels que,

$$\frac{k}{p} = \frac{\pi}{\theta_j^* - \theta_j} = \frac{k'}{p'}.$$

Ainsi a moins que $\theta_j^* - \theta_j = \pi$, on obtient une contradiction avec le fait que k et k' sont deux entiers premiers entre eux. Donc pour tout $j = 1, \dots, J$, θ_j est égal à θ_j^* (modulo $\pi\mathbb{Z}$), ce qui finit la preuve sur l'unicité du minimum de $M_1(\alpha)$. \square

Convergence uniforme de $M_{1,\epsilon}$:

On définit pour tout $(k, v) \in \mathbb{Z} \times \mathbb{R}$ les quantités suivantes,

$$\begin{aligned} c_j^\epsilon(k, v) &= a_j^{-4+\sigma-iv} e^{-ik\theta_j} \mathcal{M}_{|\hat{f}_j|^2}^\epsilon(k, v), \\ s_j^\epsilon(k, v) &= a_j^{-4+\sigma-iv} e^{-ik\theta_j} \mathcal{M}_{|\hat{w}_j|^2}^\epsilon(k, v), \\ \tilde{c}^\epsilon &= \frac{1}{J} \sum_{j=1}^J c_j^\epsilon \quad \text{et} \quad \tilde{s}^\epsilon = \frac{1}{J} \sum_{j=1}^J s_j^\epsilon. \end{aligned}$$

Le processus de contraste $M_\epsilon(\alpha_1)$ est alors la somme de trois termes :

$$M_{1,\epsilon}(\alpha) = D_{1,\epsilon}(\alpha_1) + \epsilon L_{1,\epsilon}(\alpha_1) + \epsilon^2 Q_{1,\epsilon}(\alpha_1), \quad (4.47)$$

où

$$\begin{aligned} D_{1,\epsilon}(\alpha_1) &= \frac{1}{J} \sum_{j=1}^J \int_{|v| \leq v_\epsilon} \sum_{|k| \leq k_\epsilon} |c_j^\epsilon(k, v)|^2 dv - \int_{|v| \leq v_\epsilon} \sum_{|k| \leq k_\epsilon} |\tilde{c}^\epsilon(k, v)|^2 dv \\ L_{1,\epsilon}(\alpha_1) &= \frac{2\epsilon}{J} \sum_{j=1}^J \int_{|v| \leq v_\epsilon} \sum_{|k| \leq k_\epsilon} \Re \left\{ (c_j^\epsilon(k, v) - \tilde{c}^\epsilon(k, v)) \overline{(s_j^\epsilon(k, v) - \tilde{s}^\epsilon(k, v))} \right\} dv \\ Q_{1,\epsilon}(\alpha_1) &= \frac{\epsilon^2}{J} \sum_{j=1}^J \int_{|v| \leq v_\epsilon} \sum_{|k| \leq k_\epsilon} |s_j^\epsilon(k, v)|^2 dv - \epsilon^2 \int_{|v| \leq v_\epsilon} \sum_{|k| \leq k_\epsilon} |\tilde{s}^\epsilon(k, v)|^2 dv, \end{aligned}$$

et $\Re(x)$ désigne la partie réelle du nombre complexe x . Par l'inégalité de Cauchy-Schwarz, on a :

$$|\epsilon L_{1,\epsilon}(\alpha_1)| \leq 2 \left\{ \sup_{\alpha_1 \in \tilde{\mathcal{A}}_1} |D_{1,\epsilon}(\alpha_1) - M(\alpha_1)| + \sup_{\alpha_1 \in \tilde{\mathcal{A}}_1} M(a, \theta) \right\}^{1/2} \left\{ \sup_{\alpha_1 \in \tilde{\mathcal{A}}_1} \epsilon^2 Q_{1,\epsilon}(\alpha_1) \right\}^{1/2}.$$

Puisque la fonction M_1 est continue sur l'ensemble compact $\tilde{\mathcal{A}}_1$, il suffit de montrer la convergence uniforme de $D_{1,\epsilon}$ vers M_1 et l'uniforme convergence en probabilité de $\epsilon^2 Q_{1,\epsilon}$ vers zéro.

On étudie d'abord l'uniforme convergence sur $\tilde{\mathcal{A}}_1$ en probabilité de $D_\epsilon(\alpha_1)$ vers $M(\alpha_1)$. Puisque $\tilde{\mathcal{A}}_1$ est un ensemble compact, d'après l'inégalité de Cauchy-Schwarz, il suffit de prouver que pour un $j \in \{1, \dots, J\}$ fixé le terme suivant converge vers zéro :

$$(I) = \int_{|v| \leq v_\epsilon} \sum_{|k| \leq k_\epsilon} |\mathcal{M}_{|\hat{f}_j|^2}^\epsilon(k, v)|^2 dv - \int_{|v| \leq v_\epsilon} \sum_{|k| \leq k_\epsilon} |\mathcal{M}_{|\hat{f}_j|^2}(k, v)|^2 dv$$

Comme tous nombres complexes a et b vérifient l'inégalité $|a|^2 - |b|^2 \leq |a - b|^2 + 2|(a - b)b|$, on obtient que,

$$|(I)| \leq (I - 1) + (I - 2) + (I - 3)$$

avec,

$$\begin{aligned} (I-1) &= \left\{ \int_{|v|>v_\epsilon} \sum_{|k|\leq k_\epsilon} + \int_{|v|>v_\epsilon} \sum_{|k|>k_\epsilon} + \int_{|v|\leq v_\epsilon} \sum_{|k|>k_\epsilon} \right\} |\mathcal{M}_{|\hat{f}_j|^2}(k, v)|^2 dv \\ (I-2) &= \int_{|v|\leq v_\epsilon} \sum_{|k|\leq k_\epsilon} \left| \mathcal{M}_{|\hat{f}_j|^2}^\epsilon(k, v) - \mathcal{M}_{|\hat{f}_j|^2}(k, v) \right|^2 dv \\ (I-3) &= 2 \int_{|v|\leq v_\epsilon} \sum_{|k|\leq k_\epsilon} |\mathcal{M}_{|\hat{f}_j|^2}(k, v)| \left| \mathcal{M}_{|\hat{f}_j|^2}^\epsilon(k, v) - \mathcal{M}_{|\hat{f}_j|^2}(k, v) \right| dv. \end{aligned}$$

Par l'hypothèse (4.23), le terme $(I-1)$ converge vers zéro. D'autre part, les hypothèses (4.24) et (4.25) ont pour conséquence que,

$$\left| \mathcal{M}_{|\hat{f}_j|^2}^\epsilon(k, v) - \mathcal{M}_{|\hat{f}_j|^2}(k, v) \right| \leq \int_{|\omega|>\delta_\epsilon} |\hat{f}(\omega)|^2 |\omega|^{2s} \frac{d\omega}{\delta_\epsilon^{2s+2-\sigma}}. \quad (4.48)$$

Ainsi, d'après l'inégalité de Cauchy-Schwarz et puisque δ_ϵ tend vers $+\infty$, on a,

$$(I-2) + (I-3) = o(k_\epsilon v_\epsilon \delta_\epsilon^{-4s-4+2\sigma} + k_\epsilon^{1/2} v_\epsilon^{1/2} \delta_\epsilon^{-2s-2+\sigma}).$$

L'hypothèse (4.27) permet de conclure que $D_{1,\epsilon}$ converge uniformément vers M_1 .

Maintenant, on montre l'uniforme convergence en probabilité de $\epsilon^2 Q_{1,\epsilon}$ vers zéro. Soit $j \in \{1 \dots J\}$ un entier fixé. Par l'inégalité de Cauchy-Schwarz et la compacité de l'ensemble $\tilde{\mathcal{A}}_1$, il suffit de montrer que la variable (II) converge en probabilité vers zéro :

$$(II) = \epsilon^2 \int_{|v|\leq v_\epsilon} \sum_{|k|\leq k_\epsilon} |\mathcal{M}_{w_j}^\epsilon(k, v)|^2 dv.$$

Comme tous nombres complexes a et b vérifient l'inégalité $|a+b|^2 \leq 2|a|^2 + 2|b|^2$, on obtient que :

$$(II) \leq 8(II-1) + 2(II-2),$$

avec,

$$\begin{aligned} (II-1) &= \epsilon^2 \int_{|v|\leq v_\epsilon} \sum_{|k|\leq k_\epsilon} |\mathcal{M}_{\Re \hat{f}_j * \overline{\hat{W}_j}}^\epsilon(k, v)|^2 dv, \\ (II-2) &= \epsilon^4 \int_{|v|\leq v_\epsilon} \sum_{|k|\leq k_\epsilon} |\mathcal{M}_{|\hat{W}_j|^2}^\epsilon(k, v)|^2 dv. \end{aligned}$$

On souhaite majorer l'espérance du terme $(II-1)$, on fait alors une majoration terme à terme :

$$\begin{aligned} \mathbb{E} |\mathcal{M}_{\Re \hat{f}_j * \overline{\hat{W}_j}}^\epsilon(k, v)|^2 &= \int_{\mathcal{D}} \left| \int_0^{2\pi} \int_0^{\delta_\epsilon} \left(\cos(w \cdot x) \Re \hat{f}_j(\omega) - \sin(w \cdot x) \Im \hat{f}_j(\omega) \right) e^{-ik \cdot r} r^{\sigma-iv} \frac{dr d\theta}{2\pi r} \right|^2 dx \\ &\leq d^2 \left| \int_0^{2\pi} \int_0^{\delta_\epsilon} |\hat{f}_j(\omega)| r^{\sigma-1} \frac{dr d\theta}{2\pi} \right|^2. \end{aligned}$$

Puisque $\sigma - 2 < 2s$, et par l'inégalité de Cauchy-Schwarz, on peut majorer par,

$$\mathbb{E}|\mathcal{M}_{\mathfrak{R}\hat{f}_j^* \overline{\hat{W}_j}}^\epsilon(k, v)|^2 \leq d^2 \int_{|w| \leq \delta_\epsilon} |\hat{f}_j(\omega)|^2 |\omega|^{\sigma-2} \frac{d\omega}{2\pi} \int_{|w| \leq \delta_\epsilon} |\omega|^{\sigma-2} \frac{d\omega}{2\pi}. \quad (4.49)$$

Par conséquent, on a : $\mathbb{E}|\mathcal{M}_{\mathfrak{R}\hat{f}_j^* \overline{\hat{W}_j}}^\epsilon(k, v)|^2 = \mathcal{O}(\delta_\epsilon^\sigma)$. De la même manière, on obtient :

$$\mathbb{E}|\mathcal{M}_{|\hat{W}_j|^2}^\epsilon(k, v)|^2 = \mathcal{O}(\delta_\epsilon^{2\sigma}). \quad (4.50)$$

Ainsi, l'hypothèse (4.28) implique la convergence en probabilité de (II), car :

$$\mathbb{E}(II) = \mathcal{O}(\epsilon^2 k_\epsilon v_\epsilon \delta_\epsilon^\sigma (1 + \epsilon^2 \delta_\epsilon^\sigma)).$$

Donc $\epsilon^2 Q_{1,\epsilon}$ converge uniformément en probabilité vers zéro. \square

Preuve du théorème 4.2

La preuve de ce théorème est identique dans les grandes lignes à celle du théorème 2.2. Puisque $\alpha_{1,\epsilon}$ est défini comme le minimum du processus de contraste $M_{1,\epsilon}(\cdot)$, on a :

$$\nabla M_{1,\epsilon}(\hat{\alpha}_{1,\epsilon}) = 0,$$

où ∇ est l'opérateur gradient. Par un développement de Taylor, il existe $\bar{\alpha}_{1,\epsilon}$ au voisinage de α_1^* de telle sorte que

$$\nabla^2 M_{1,\epsilon}(\bar{\alpha}_{1,\epsilon}) \epsilon^{-1} (\hat{\alpha}_{1,\epsilon} - \alpha_1^*) = -\epsilon^{-1} \nabla M_{1,\epsilon}(\alpha_1^*),$$

où ∇^2 est l'opérateur hessien. La proposition 4.1 montre que la variable $\epsilon^{-1} \nabla M_{1,\epsilon}(\alpha_1^*)$ converge en loi vers un vecteur gaussien, et la proposition 4.2 montre que la matrice hessienne $\nabla^2 M_{1,\epsilon}(\bar{\alpha}_{1,\epsilon})$ converge en probabilité vers une matrice inversible H . Ensuite par le lemme de Slutsky, le théorème suit. \square

Les propositions 4.1 et 4.2 nécessitent le calcul des dérivées premières et secondes de M_ϵ . Ces calculs sont présentés succinctement dans le lemme suivant :

Lemme 4.1. *Soient $2 \leq j, l \leq J$, $j \neq l$ des entiers. Pour $\alpha \in \mathcal{A}_1$, les dérivées partielles du premier ordre premières de M_ϵ sont,*

$$\begin{aligned} \frac{d}{da_j} M_\epsilon(\alpha) &= \frac{2\Re}{J} \int_{|v| \leq v_\epsilon} \sum_{|k| \leq k_\epsilon} \frac{-4 + \sigma - iv}{a_j} \left\{ |d_j^\epsilon(k, v)|^2 - d_j^\epsilon(k, v) \overline{\tilde{d}^\epsilon(k, v)} \right\} dv, \\ \frac{d}{d\theta_j} M_\epsilon(\alpha) &= \frac{2\Re}{J} \int_{|v| \leq v_\epsilon} \sum_{|k| \leq k_\epsilon} \left\{ i k d_j^\epsilon(k, v) \overline{\tilde{d}^\epsilon(k, v)} \right\} dv, \end{aligned}$$

et celles du deuxième ordre sont,

$$\begin{aligned}
\frac{d^2}{da_j^2} M_\epsilon(\alpha) &= \frac{2\Re}{J} \int_{-v_\epsilon}^{v_\epsilon} \sum_{|k| \leq k_\epsilon} \frac{(-4 + \sigma - iv)(-5 + \sigma - iv)}{a_j^2} \left\{ 2|d_j^\epsilon(k, v)|^2 - d_j^\epsilon(k, v) \overline{\tilde{d}^\epsilon(k, v)} \right\} dv \\
&\quad - \frac{2\Re}{J} \int_{-v_\epsilon}^{v_\epsilon} \sum_{|k| \leq k_\epsilon} \frac{|-4 + \sigma - iv|^2}{a_j^2} 2|d_j^\epsilon(k, v)|^2 dv, \\
\frac{d^2}{da_j da_l} M_\epsilon(\alpha) &= \frac{-2\Re}{J^2} \int_{-v_\epsilon}^{v_\epsilon} \sum_{|k| \leq k_\epsilon} \frac{|-4 + \sigma - iv|^2}{a_j a_l} d_j^\epsilon(k, v) \overline{\tilde{d}_l^\epsilon(k, v)} dv, \\
\frac{d^2}{d\theta_j^2} M_\epsilon(\alpha) &= \frac{2\Re}{J} \int_{-v_\epsilon}^{v_\epsilon} \sum_{|k| \leq k_\epsilon} k^2 \left\{ d_j^\epsilon(k, v) \overline{\tilde{d}^\epsilon(k, v)} - |d_j^\epsilon(k, v)|^2 / J \right\} dv, \\
\frac{d^2}{d\theta_j d\theta_l} M_\epsilon(\alpha) &= \frac{-2\Re}{J^2} \int_{-v_\epsilon}^{v_\epsilon} \sum_{|k| \leq k_\epsilon} k^2 d_j^\epsilon(k, v) \overline{\tilde{d}_l^\epsilon(k, v)} dv, \\
\frac{d^2}{d\theta_j a_j} M_\epsilon(\alpha) &= \frac{2\Re}{J} \int_{-v_\epsilon}^{v_\epsilon} \sum_{|k| \leq k_\epsilon} \frac{k}{a_j} \left\{ i \frac{-4 + \sigma + iv}{J} |d_j^\epsilon(k, v)|^2 + i(4 - \sigma + iv) d_j^\epsilon(k, v) \overline{\tilde{d}^\epsilon(k, v)} \right\} dv, \\
\frac{d^2}{d\theta_j a_l} M_\epsilon(\alpha) &= \frac{2\Re}{J^2} \int_{-v_\epsilon}^{v_\epsilon} \sum_{|k| \leq k_\epsilon} i k \frac{-4 + \sigma + iv}{a_l} d_j^\epsilon(k, v) \overline{\tilde{d}^\epsilon(k, v)} dv.
\end{aligned}$$

Proposition 4.1. *Sous les hypothèses du théorème 4.2, la variable $\epsilon^{-1} \nabla M_{1,\epsilon}(\alpha_1^*)$ converge en loi vers un vecteur gaussien centré.*

Preuve de la proposition 4.1. On reprend la décomposition (4.47) du processus de contraste. On souhaite d'abord montrer que

$$\epsilon^{-1} \nabla M_{1,\epsilon}(\alpha_1^*) = \nabla L_{1,\epsilon}(\alpha_1^*) + o_{\mathbb{P}}(1).$$

Ensuite, on prouve que le vecteur $\epsilon \nabla L_\epsilon(\alpha_1^*)$ converge en loi vers un vecteur gaussien centré.

On montre d'abord que $\epsilon^{-1} \nabla D_\epsilon(\alpha_1^*)$ converge vers zéro. D'après le lemme 4.1, et après quelques calculs, nous avons pour $k = 2 \dots J$:

$$\begin{aligned}
\frac{d}{da_j} D_{1,\epsilon}(\alpha_1^*) &= \frac{2\Re}{J} \int_{|v| \leq v_\epsilon} \sum_{|k| \leq k_\epsilon} \frac{-4 + \sigma - iv}{a_j^*} \left\{ |c_j^\epsilon(k, v)|^2 - c_j^\epsilon(k, v) \overline{\tilde{c}^\epsilon(k, v)} \right\} dv, \\
\frac{d}{d\theta_j} D_{1,\epsilon}(\alpha_1^*) &= \frac{2\Re}{J} \int_{|v| \leq v_\epsilon} \sum_{|k| \leq k_\epsilon} \left\{ i k c_j^\epsilon(k, v) \overline{\tilde{c}^\epsilon(k, v)} \right\} dv.
\end{aligned}$$

Or, par définition, les égalités suivantes sont vraies :

$$\begin{aligned}
0 &= \Re \int_{|v| \leq v_\epsilon} \sum_{|k| \leq k_\epsilon} \frac{-iv}{a_j^*} \left\{ |\mathcal{M}_{|\hat{f}|^2}(k, v)|^2 - \mathcal{M}_{|\hat{f}|^2}(k, v) \overline{\mathcal{M}_{|\hat{f}|^2}(k, v)} \right\} dv, \\
0 &= \Re \int_{|v| \leq v_\epsilon} \sum_{|k| \leq k_\epsilon} \left\{ i k \mathcal{M}_{|\hat{f}|^2}(k, v) \overline{\mathcal{M}_{|\hat{f}|^2}(k, v)} \right\} dv.
\end{aligned}$$

On obtient alors que :

$$\begin{aligned} \frac{d}{da_j} D_{1,\epsilon}(\alpha_1^*) &= \mathcal{O} \left\{ \int_{|v| \leq v_\epsilon} \sum_{|k| \leq k_\epsilon} |v| \left| \mathcal{M}_{|\hat{f}|^2}^\epsilon(k, v) - \mathcal{M}_{|\hat{f}|^2}(k, v) \right|^2 dv \right\} \\ &\quad + \mathcal{O} \left\{ \int_{|v| \leq v_\epsilon} \sum_{|k| \leq k_\epsilon} |v| |\mathcal{M}_{|\hat{f}|^2}(k, v)| \left| \mathcal{M}_{|\hat{f}|^2}^\epsilon(k, v) - \mathcal{M}_{|\hat{f}|^2}(k, v) \right|^2 dv \right\}, \\ \frac{d}{d\theta_j} D_{1,\epsilon}(\alpha_1^*) &= \mathcal{O} \left\{ \int_{|v| \leq v_\epsilon} \sum_{|k| \leq k_\epsilon} |k| \left| \mathcal{M}_{|\hat{f}|^2}^\epsilon(k, v) - \mathcal{M}_{|\hat{f}|^2}(k, v) \right|^2 dv \right\} \\ &\quad + \mathcal{O} \left\{ \int_{|v| \leq v_\epsilon} \sum_{|k| \leq k_\epsilon} |k| |\mathcal{M}_{|\hat{f}|^2}(k, v)| \left| \mathcal{M}_{|\hat{f}|^2}^\epsilon(k, v) - \mathcal{M}_{|\hat{f}|^2}(k, v) \right|^2 dv \right\}. \end{aligned}$$

D'après les hypothèses (4.31) et (4.32), l'inégalité (4.48) est donc vérifiée, et nous obtenons que :

$$\begin{aligned} \frac{d}{da_j} D_\epsilon(\alpha_1^*) &= o \left(v_\epsilon^2 k_\epsilon \delta_\epsilon^{-4s-4+2\sigma} + \delta_\epsilon^{-2s-2+1\sigma} \right), \\ \frac{d}{d\theta_j} D_\epsilon(\alpha_1^*) &= o \left(v_\epsilon k_\epsilon^2 \delta_\epsilon^{-4s-4+2\sigma} + \delta_\epsilon^{-2s-2+1\sigma} \right). \end{aligned}$$

Par l'hypothèse (4.35), $\epsilon^{-1} \nabla D_{1,\epsilon}(\alpha_1^*)$ converge vers zéro.

Étudions maintenant la convergence en probabilité de $\epsilon \nabla Q_\epsilon(\alpha_1^*)$ vers zéro. D'après le lemme 4.1, nous avons pour $k = 2 \dots J$,

$$\epsilon \frac{d}{da_j} Q_{1,\epsilon}(\alpha_1^*) = \epsilon \frac{2\Re}{J^2} \sum_{\substack{l=1 \dots J \\ l \neq j}} \int_{|v| \leq v_\epsilon} \sum_{|k| \leq k_\epsilon} \frac{-4 + \sigma - iv}{a_j^*} s_j^\epsilon(k, v) \overline{\tilde{s}_l^\epsilon(k, v)} dv \quad (4.51)$$

$$+ \epsilon \frac{2\Re}{J} \left(1 - \frac{1}{J}\right) \int_{|v| \leq v_\epsilon} \sum_{|k| \leq k_\epsilon} \frac{-4 + \sigma - iv}{a_j^*} |s_j^\epsilon(k, v)|^2 dv, \quad (4.52)$$

$$\epsilon \frac{d}{d\theta_j} Q_{1,\epsilon}(\alpha_1^*) = \epsilon \frac{2\Re}{J^2} \sum_{\substack{l=1 \dots J \\ l \neq j}} \int_{|v| \leq v_\epsilon} \sum_{|k| \leq k_\epsilon} \left\{ i k s_j^\epsilon(k, v) \overline{\tilde{s}_l^\epsilon(k, v)} \right\} dv. \quad (4.53)$$

En utilisant les calculs (4.49) et (4.50) de la preuve du théorème 4.1, nous avons que :

$$(4.52) = \mathcal{O}_{\mathbb{P}} \left(\epsilon k_\epsilon v_\epsilon \delta_\epsilon^\sigma (1 + \epsilon^2 \delta_\epsilon^\sigma) \right).$$

Et donc (4.52) converge en probabilité vers zéro par l'hypothèse (4.36). D'autre part, en utilisant (4.49) et (4.50), on a :

$$\mathbb{E}|(4.51)| = \mathcal{O} \left(\epsilon k_\epsilon v_\epsilon^2 \delta_\epsilon^\sigma (1 + \epsilon^2 \delta_\epsilon^\sigma) \right),$$

$$\mathbb{E}|(4.53)| = \mathcal{O} \left(\epsilon k_\epsilon^2 v_\epsilon \delta_\epsilon^\sigma (1 + \epsilon^2 \delta_\epsilon^\sigma) \right).$$

Par conséquent $\epsilon \nabla Q_\epsilon$ converge en probabilité vers zéro d'après l'hypothèse (4.36). De ce fait on a bien que :

$$\epsilon^{-1} \nabla M_{1,\epsilon}(\alpha_1^*) = \nabla L_{1,\epsilon}(\alpha_1^*) + o_{\mathbb{P}}(1).$$

D'après le lemme 4.1, on a pour $k = 2 \dots J$:

$$\begin{aligned} \frac{d}{da_j} L_{1,\epsilon}(\alpha_1^*) &= \frac{2\Re}{J} \int \sum_{|v| \leq v_\epsilon, |k| \leq k_\epsilon} \frac{-4 + \sigma - iv}{a_j^*} \left\{ c_j^\epsilon(k, v) \overline{(s_j^\epsilon - \tilde{s}^\epsilon)(k, v)} - (c_j^\epsilon - \tilde{c}^\epsilon)(k, v) \overline{s_j^\epsilon(k, v)} \right\} dv, \\ \frac{d}{d\theta_j} L_{1,\epsilon}(\alpha_1^*) &= \frac{2\Re}{J} \int \sum_{|v| \leq v_\epsilon, |k| \leq k_\epsilon} ik \left\{ c_j^\epsilon(k, v) \overline{\tilde{s}^\epsilon(k, v)} - \tilde{c}_j^\epsilon(k, v) \overline{s_j^\epsilon(k, v)} \right\} dv. \end{aligned}$$

D'après (4.48), on peut écrire que :

$$c_j^\epsilon(k, v) = \mathcal{M}_{|\hat{f}|^2}(k, v) + o(\delta_\epsilon^{-2s-2+\sigma}) \quad \forall j = \dots J.$$

De plus, par (4.49) et (4.50), la propriété suivante est vérifiée :

$$\mathbb{E}|s_j^\epsilon(k, v)| = \mathcal{O}(\delta_\epsilon^{\sigma/2} + \epsilon \delta_\epsilon^\sigma) \quad \forall j = \dots J.$$

Ces dernières égalités permettent de récrire les dérivées partielles de $L_{1,\epsilon}$ comme :

$$\begin{aligned} \frac{d}{da_j} L_{1,\epsilon}(\alpha_1^*) &= \frac{2\Re}{J} \int \sum_{|v| \leq v_\epsilon, |k| \leq k_\epsilon} \frac{-4 + \sigma - iv}{a_j^*} \left\{ \mathcal{M}_{|\hat{f}|^2}(k, v) \overline{(s_j^\epsilon - \tilde{s}^\epsilon)(k, v)} \right\} dv \\ &\quad + o_{\mathbb{P}}(k_\epsilon v_\epsilon^2 \delta_\epsilon^{-2s-2+3\sigma/2} (1 + \epsilon \delta_\epsilon^{\sigma/2})), \\ \frac{d}{d\theta_j} L_{1,\epsilon}(\alpha_1^*) &= \frac{2\Re}{J} \int \sum_{|v| \leq v_\epsilon, |k| \leq k_\epsilon} ik \left\{ \mathcal{M}_{|\hat{f}|^2}(k, v) \overline{\tilde{s}^\epsilon(k, v)} - \mathcal{M}_{|\hat{f}|^2}(k, v) \overline{s_j^\epsilon(k, v)} \right\} dv, \\ &\quad + o_{\mathbb{P}}(k_\epsilon^2 v_\epsilon \delta_\epsilon^{-2s-2+3\sigma/2} (1 + \epsilon \delta_\epsilon^{\sigma/2})). \end{aligned}$$

Puisque $\mathbb{E}|\mathcal{M}_{|\hat{W}_j(\omega)|^2}^\epsilon(k, v)| = \mathcal{O}(\delta_\epsilon^\sigma)$, on a alors que

$$s_j^\epsilon(k, v) = a_j^{-4+\sigma-iv} e^{-ik\theta_j} \mathcal{M}_{2\Re\{\hat{f}_j \overline{\hat{W}_j}\}}^\epsilon(k, v) + \mathcal{O}_{\mathbb{P}}(\epsilon \delta_\epsilon^\sigma).$$

Par conséquent, les hypothèses (4.30) et (4.36)-(4.37) impliquent que :

$$\begin{aligned} \frac{d}{da_j} L_{1,\epsilon}(\alpha_1^*) &= \frac{2\Re}{J} (V_{a,j} - \frac{1}{J} \sum_{l=1}^J V_{a,l}) + o_{\mathbb{P}}(1), \\ \frac{d}{d\theta_j} L_{1,\epsilon}(\alpha_1^*) &= \frac{2\Re}{J} (\frac{1}{J} \sum_{l=1}^J V_{\theta,l} - V_{\theta,j}) + o_{\mathbb{P}}(1), \end{aligned}$$

où

$$V_{a,l} = \Re \int \sum_{|v| \leq v_\epsilon, |k| \leq k_\epsilon} \frac{-4 + \sigma - iv}{a_j^*} \left\{ \mathcal{M}_{|\hat{f}|^2}(k, v) a_l^{*-4 + \sigma + iv} e^{ik\theta_l^*} \overline{\mathcal{M}_{2\Re\{\hat{f}_l \bar{W}_l\}}(k, v)} \right\} dv$$

$$V_{\theta,l} = \Re \int \sum_{|v| \leq v_\epsilon, |k| \leq k_\epsilon} ik \mathcal{M}_{|\hat{f}|^2}(k, v) a_l^{*-4 + \sigma + iv} e^{ik\theta_l^*} \overline{\mathcal{M}_{2\Re\{\hat{f}_l \bar{W}_l\}}(k, v)} dv.$$

Il suffit alors d'étudier le comportement asymptotique des deux variables centrées $V_{a,l}$ et $V_{\theta,l}$ (pour un $l \in \{1 \dots J\}$ fixé).

Posons $g(r, \theta)$ la fonction suivante $r^\sigma |\hat{f}(r, \theta)|^2$, et $h(r, \theta)$ la fonction $r^\sigma \hat{f}(r, \theta)$. D'après l'hypothèse (4.32), ces fonctions appartiennent à $\mathbb{L}^2(\mathcal{G})$. Soient $\partial_r g$ et $\partial_\theta g$ les dérivées partielles de g respectivement par rapport aux variables r et θ . Après quelques calculs utilisant la transformée de Fourier Mellin inverse (4.9), nous avons que :

$$V_{a,l} = \int_{\mathcal{D}} dW_l(x) \frac{2}{a_j^*} \left\langle (-4 + \sigma)g - \partial_r g, \Re \left(h e^{i\omega \cdot \frac{1}{a_l^*} A_{\theta_l^*} x} \right) \right\rangle_{L^2(\mathcal{G})} + o_{\mathbb{P}}(1),$$

$$V_{\theta,l} = \int_{\mathcal{D}} dW_l(x) 2 \left\langle \partial_\theta g, \Re \left(h e^{i\omega \cdot \frac{1}{a_l^*} A_{\theta_l^*} x} \right) \right\rangle_{L^2(\mathcal{G})} + o_{\mathbb{P}}(1).$$

Par conséquent les variances des dérivées partielles de $M_{1,\epsilon}$ convergent respectivement vers,

$$\begin{aligned} \text{var} \left(\frac{d}{da_j} L_{1,\epsilon}(\alpha^*) \right) &= \frac{16}{a_j^{*2} J^4} \sum_{l \neq j} \int_{\mathcal{D}} \left\langle (-4 + \sigma)g - \partial_r g, \Re \left(h e^{i\omega \cdot \frac{1}{a_l^*} A_{\theta_l^*} x} \right) \right\rangle_{L^2(\mathcal{G})}^2 dx \\ &\quad + \frac{16(J-1)^2}{a_j^{*2} J^4} \int_{\mathcal{D}} \left\langle (-4 + \sigma)g - \partial_r g, \Re \left(h e^{i\omega \cdot \frac{1}{a_j^*} A_{\theta_j^*} x} \right) \right\rangle_{L^2(\mathcal{G})}^2 dx, \\ \text{var} \left(\frac{d}{d\theta_j} L_{1,\epsilon}(\alpha^*) \right) &= \frac{16}{J^4} \sum_{l \neq j} \int_{\mathcal{D}} \left\langle \partial_\theta g, \Re \left(h e^{i\omega \cdot \frac{1}{a_l^*} A_{\theta_l^*} x} \right) \right\rangle_{L^2(\mathcal{G})}^2 dx \\ &\quad + \frac{16(J-1)^2}{J^4} \int_{\mathcal{D}} \left\langle \partial_\theta g, \Re \left(h e^{i\omega \cdot \frac{1}{a_j^*} A_{\theta_j^*} x} \right) \right\rangle_{L^2(\mathcal{G})}^2 dx. \end{aligned}$$

□

Proposition 4.2. Posons $\mathcal{A}_{1,\epsilon}^{loc}$ le sous-ensemble de \mathcal{A}_1 suivant :

$$\mathcal{A}_{1,\epsilon}^{loc} = \left\{ \alpha_1 \in \tilde{\mathcal{A}}_1, \|\alpha_1 - \alpha_1^*\| \leq \|\hat{\alpha}_{1,\epsilon} - \alpha_1^*\| \right\}.$$

Sous les hypothèses du théorème 4.2, il existe une matrice inversible H telle que

$$\sup_{(\bar{\alpha}_{1,\epsilon}) \in \mathcal{A}_{1,\epsilon}} \|\nabla^2 M_{1,\epsilon}(\bar{\alpha}_{1,\epsilon}) - H\| = o_{\mathbb{P}}(1)$$

Preuve de la proposition 4.2. La matrice limite est la valeur de la matrice hessienne de M_1 au point α_1^* .

De la même façon que dans la preuve du théorème 4.1, on prouve que $\nabla^2 M_{1,\epsilon}$ converge uniformément vers $\nabla^2 M_1$ sur un voisinage compact de α_1^* . Grâce à la propriété de continuité uniforme de $\nabla^2 M_1$ sur l'espace compact $\tilde{\mathcal{A}}_1$, on prouve avec les mêmes arguments développés dans la preuve de la proposition 2.6 que

$$\sup_{\bar{\alpha}_{1,\epsilon} \in \mathcal{A}_{1,\epsilon}} \|\nabla^2 M_{1,\epsilon}(\bar{\alpha}_{1,\epsilon}) - H\| = o_{\mathbb{P}}(1).$$

Montrons maintenant que la matrice H est inversible. Après calculs, la matrice H est diagonale par bloc :

$$H = \frac{2}{J} \begin{pmatrix} H_a & 0 \\ 0 & H_\theta \end{pmatrix},$$

où H_a et H_θ sont définies de la façon suivante :

$$\begin{aligned} H_a &= ((\sigma - 5)D_a^{-2} - \frac{\sigma - 4}{J}a_{-1}a_{-1}^T)(\sigma - 4)\|g\|^2 + (D_a^2 - \frac{1}{J}a_{-1}a_{-1}^T)\|\partial_r g\|^2, \\ H_\theta &= (I_{J-1} - \frac{1}{J}\mathbb{I}_{J-1}\mathbb{I}_{J-1}^T)\|\partial_\theta g\|^2. \end{aligned}$$

$D_a = \text{diag}(a_2^* \dots a_J^*)$, $I_{J-1} = \text{diag}(1 \dots 1)$ sont des matrices diagonales de taille $J - 1$. $a_{-1} = D_a^{-1}\mathbb{I}_{J-1}$ et $\mathbb{I}_{J-1} = (1, \dots, 1)^T$ sont des vecteurs de \mathbb{R}^{J-1} .

Si $\partial_\theta g$ n'est pas identiquement nulle, la matrice H_θ est inversible,

$$H_\theta^{-1} = (I_{J-1} + \mathbb{I}_{J-1})\|\partial_\theta g\|_{L^2(\mathcal{G})}^{-2}.$$

Par ailleurs, on récrit la matrice H_a comme,

$$H_a = D_a^{-2}\gamma_0 - a_{-1}a_{-1}^T\gamma_1,$$

où $\gamma_0 = (\sigma - 5)(\sigma - 4)\|g\|_{L^2(\mathcal{G})}^2 + \|\partial_r g\|_{L^2(\mathcal{G})}^2$ et $\gamma_1 = (\frac{(\sigma-4)^2}{J}\|g\|_{L^2(\mathcal{G})}^2 + \frac{1}{J}\|\partial_r g\|_{L^2(\mathcal{G})}^2)$. Cette matrice est inversible si, et seulement si,

$$\gamma_0 \neq 0 \quad \text{et} \quad \gamma_1(J - 1) - \gamma_0 \neq 0.$$

Ceci se traduit par l'hypothèse (4.33), et donc la matrice inverse de H_a est égale à,

$$H_a^{-1} = \gamma_0^{-1}D_a^2 - \frac{\gamma_1\gamma_0^{-1}}{\gamma_0 - \gamma_1(J - 1)}aa^T,$$

où $a = D_a\mathbb{I}_{J-1}$. □

Preuve du théorème 4.3

La preuve de ce théorème suit les lignes classiques de la convergence des M -estimateurs. On montre que $M_2(\alpha_2)$ admet un unique minimum en $\alpha_2 = \alpha_2^*$, et que $M_{2,\epsilon}$ converge uniformément (en α_2) en probabilité vers M_2 i.e.

$$\sup_{\alpha_2 \in \mathcal{A}_2} |M_{2,\epsilon}(\alpha_2) - M_2(\alpha_2)| = o_{\mathbb{P}}(1).$$

Ces deux conditions assurent que $\hat{\alpha}_{2,\epsilon}$ converge en probabilité vers α_2^* quand $\epsilon \rightarrow 0$ (voir le chapitre 1).

Unicité du minimum de M_2 :

On rappelle que

$$M_2(\alpha_2) = \int_{\mathbb{R}^2} |\hat{f}|^2(\omega) \left(1 - \left| \frac{1}{J} \sum_{j'=1}^J e^{i\omega \cdot (\tilde{\beta}_{j'}^* - \tilde{\beta}_{j'}^*)} \right|^2 \right) d\omega.$$

Ainsi M_2 a un minimum en α_2^* car $M_2(\alpha_2^*) = 0$. Supposons qu'il existe $\alpha_2 \in \mathcal{A}_2$ tel que $M_2(\alpha_2) = 0$. Puisque \hat{f} est une fonction continue et non nulle, il existe un ensemble ouvert U de \mathbb{R}^2 tel que $U \subset \text{supp}(\hat{f})$. On a alors,

$$1 = \left| \frac{1}{J} \sum_{j=1}^J \sum_{j'=1}^J e^{i\omega \cdot (\tilde{\beta}_{j'}^* - \tilde{\beta}_{j'}^*)} \right|^2, \quad \forall \omega \in U.$$

Cette égalité est le cas d'égalité de l'inégalité de Cauchy-Schwarz. Elle est vérifiée si, et seulement si pour tout $\omega \in U$ il existe $\lambda_\omega \in \mathbb{R}$ tel que

$$\forall j = 1 \dots J, \quad e^{i\omega \cdot (\tilde{\beta}_j^* - \tilde{\beta}_j^*)} = \lambda_\omega.$$

Or d'après la contrainte d'identifiabilité ($\tilde{\beta}_1^* - \tilde{\beta}_1^* = 0$), λ_ω est égal à 1 pour tout $\omega \in U$. Par conséquent, nous avons que,

$$\forall \omega \in U, \quad \forall j = 1 \dots J, \quad \omega \cdot (\tilde{\beta}_j^* - \tilde{\beta}_j^*) \equiv 0 \pmod{2\pi}. \quad (4.54)$$

Supposons, par l'absurde, qu'il existe j tel que $\tilde{\beta}_j^* - \tilde{\beta}_j^* \neq 0$. Puisque l'ensemble des nombres rationnels et l'ensemble des nombres irrationnels sont deux sous-espaces denses de \mathbb{R} , il existe $\omega \in \mathbb{Q}^2 \cap U$ et $\lambda \in \mathbb{R} \setminus \mathbb{Q}$ tel que $\omega \neq 0$ et $\lambda\omega \in U$. Par l'équation (4.54), il existe $(k, k') \in \mathbb{Z}^2$ tel que

$$\omega \cdot (\tilde{\beta}_j^* - \tilde{\beta}_j^*) = 2\pi k \quad \text{et} \quad \lambda\omega \cdot (\tilde{\beta}_j^* - \tilde{\beta}_j^*) = 2\pi k'.$$

Ce qui est impossible. On en déduit que $\alpha_2 = \alpha_2^*$. □

Convergence uniforme de $M_{2,\epsilon}$:

On récrit le processus de contraste $M_{2,\epsilon}$ sous la forme suivante pour $\alpha_2 \in \mathcal{A}_2$,

$$M_{2,\epsilon}(\alpha_2) = \frac{1}{J} \sum_{j=1}^J \int_{|\omega| \leq \omega_\epsilon} |\hat{Z}_j(\omega)|^2 d\omega - \int_{|\omega| \leq \omega_\epsilon} \left| \frac{1}{J} \sum_{j=1}^J \hat{Z}_j(\omega) \right|^2 d\omega.$$

Puisque le premier terme est une constante ne dépendant pas de la variable α_2 , il suffit d'étudier la convergence uniforme du second terme. Ce dernier peut se récrire comme la somme de trois termes,

$$\int_{|\omega| \leq \omega_\epsilon} \left| \frac{1}{J} \sum_{j=1}^J \hat{Z}_j(\omega) \right|^2 d\omega = D_{2,\epsilon}(\alpha_2) + \epsilon L_{2,\epsilon}(\alpha_2) + \epsilon^2 Q_{2,\epsilon}(\alpha_2), \quad (4.55)$$

où

$$\begin{aligned} D_{2,\epsilon}(\alpha_2) &= \int_{|\omega| \leq \omega_\epsilon} \left| \frac{1}{J} \sum_{j=1}^J \left(\frac{a_j^*}{\hat{a}_{j,\epsilon}} \right)^2 \hat{f} \left(\frac{a_j^*}{\hat{a}_{j,\epsilon}} A_{\theta_j^* - \hat{\theta}_{j,\epsilon}} \omega \right) e^{i\omega \cdot (\tilde{\beta}_j - \tilde{\beta}_j^*)} \right|^2 d\omega, \\ L_{2,\epsilon}(\alpha_2) &= 2\Re \int_{|\omega| \leq \omega_\epsilon} \left[\frac{1}{J} \sum_{j=1}^J \left(\frac{a_j^*}{\hat{a}_{j,\epsilon}} \right)^2 \hat{f} \left(\frac{a_j^*}{\hat{a}_{j,\epsilon}} A_{\theta_j^* - \hat{\theta}_{j,\epsilon}} \omega \right) e^{i\omega \cdot (\tilde{\beta}_j - \tilde{\beta}_j^*)} \right] \\ &\quad \times \left[\frac{1}{J} \sum_{j=1}^J \frac{e^{-i\omega \cdot \tilde{\beta}_j}}{\hat{a}_{j,\epsilon}^2} \overline{\hat{W} \left(\frac{1}{\hat{a}_{j,\epsilon}} \hat{A}_{-\hat{\theta}_{j,\epsilon}} \omega \right)} \right] d\omega, \\ Q_{2,\epsilon}(\alpha_2) &= \int_{|\omega| \leq \omega_\epsilon} \left| \frac{1}{J} \sum_{j=1}^J \frac{e^{i\omega \cdot \tilde{\beta}_j}}{\hat{a}_{j,\epsilon}^2} \hat{W} \left(\frac{1}{\hat{a}_{j,\epsilon}} \hat{A}_{-\hat{\theta}_{j,\epsilon}} \omega \right) \right|^2 d\omega. \end{aligned}$$

Comme pour la preuve du théorème 4.1, la convergence uniforme se déduit de l'uniforme convergence en probabilité de $D_{2,\epsilon}$, et de l'uniforme convergence en probabilité de $\epsilon^2 Q_{2,\epsilon}$ vers zéro.

D'abord, étudions la convergence de $D_{2,\epsilon}$ vers la fonction :

$$\int_{\mathbb{R}^2} |\hat{f}(\omega)|^2 \left| \frac{1}{J} \sum_{j=1}^J e^{i\omega \cdot (\beta_j - \beta_j^*)} \right|^2 d\omega.$$

La différence $D_{2,\epsilon}$ avec sa limite peut se récrire comme la somme de deux termes,

$$\left(D_{2,\epsilon}(\alpha_2) - \int_{|\omega| \leq \omega_\epsilon} |\hat{f}(\omega)|^2 \left| \frac{1}{J} \sum_{j=1}^J e^{i\omega \cdot (\beta_j - \beta_j^*)} \right|^2 d\omega \right) - \int_{|\omega| > \omega_\epsilon} |\hat{f}(\omega)|^2 \left| \frac{1}{J} \sum_{j=1}^J e^{i\omega \cdot (\beta_j - \beta_j^*)} \right|^2 d\omega.$$

D'après l'hypothèse (4.31), le second terme, qui est déterministe, converge vers zéro. En utilisant l'inégalité de Cauchy-Schwarz et l'inégalité $|a|^2 - |b|^2| \leq |a - b| + 2|b||a - b|$

vérifiée pour tout $(a, b) \in \mathbb{C}^2$, le premier terme converge uniformément en probabilité vers zéro si, et seulement si pour tout $j = 1, \dots, J$ les termes suivants convergent en probabilité vers zéro,

$$(III)_j = \int_{|\omega| \leq \omega_\epsilon} \left| \left(\frac{a_j^*}{\hat{a}_{j,\epsilon}} \right)^2 \hat{f} \left(\frac{a_j^*}{\hat{a}_{j,\epsilon}} A_{\theta_j^* - \hat{\theta}_{j,\epsilon}} \omega \right) e^{i\omega \cdot (\tilde{\beta}_j - \tilde{\beta}_j^*)} - \hat{f}(\omega) e^{i\omega \cdot (\beta_j - \beta_j^*)} \right|^2 d\omega.$$

Puisque $\hat{a}_{j,\epsilon}$ et $\hat{\theta}_{j,\epsilon}$ sont des estimateurs ϵ^{-1} -consistants de a_j^* et θ_j^* , la méthode Delta implique que,

$$\left(\frac{a_j^*}{\hat{a}_{j,\epsilon}} \right)^2 - 1 = \mathcal{O}_{\mathbb{P}}(\epsilon) \quad \text{et} \quad \frac{a_j^*}{\hat{a}_{j,\epsilon}} A_{\theta_j^* - \hat{\theta}_{j,\epsilon}} - I_2 = \mathcal{O}_{\mathbb{P}}(\epsilon).$$

Par ailleurs si la fonction \hat{f} est une fonction Lipschitz, nous avons alors :

$$(III)_j = \mathcal{O}_{\mathbb{P}} \left(\int_{|\omega| \leq \omega_\epsilon} \epsilon^2 |\omega|^2 + \epsilon |\hat{f}|^2(\omega) + \epsilon^2 |\omega|^2 |\hat{f}|^2(\omega) d\omega \right).$$

Par conséquent $(III)_j$ converge uniformément vers zéro par l'hypothèse (4.41). Et donc $D_{2,\epsilon}$ converge uniformément en probabilité.

Étudions maintenant la convergence de $\epsilon^2 Q_{2,\epsilon}$. Par l'inégalité de Cauchy-Schwarz, $\epsilon^2 Q_{2,\epsilon}$ est uniformément borné par les variables suivantes,

$$(IV)_j = \epsilon^2 \int_{|\omega| \leq \omega_\epsilon} \frac{1}{\hat{a}_{j,\epsilon}^4} \left| \hat{W}_j \left(\frac{1}{\hat{a}_{j,\epsilon}} A_{-\hat{\theta}_{j,\epsilon}} \omega \right) \right| d\omega \quad j = 1, \dots, J.$$

Puisque $\hat{a}_{j,\epsilon}$ sont des estimateurs consistants et par un changement de variable, nous avons pour un $j = 1, \dots, J$ fixé que

$$(IV)_j = \mathcal{O}(\epsilon^2) \int_{|\omega| \leq \omega_\epsilon / a_{\min}} \left| \hat{W}_j(\omega) \right| d\omega.$$

Par conséquent, $\epsilon^2 Q_{2,\epsilon}$ converge uniformément vers zéro. \square

Preuve du théorème 4.4

Par les mêmes arguments que pour la preuve du théorème 4.2, il suffit de montrer que la variable $\epsilon^{-1} \nabla M_{2,\epsilon}(\alpha_2^*)$ converge en loi vers une variable gaussienne centrée (proposition 4.3), et que la matrice hessienne $\nabla^2 M_{2,\epsilon}(\bar{\alpha}_{2,\epsilon})$ converge en probabilité vers une matrice inversible H_b (proposition 4.4). \square

Les propositions 4.3 et 4.4 nécessitent le calcul des dérivées premières et secondes de $M_{2,\epsilon}$. Ces calculs sont présentés succinctement dans le lemme suivant :

Lemme 4.2. Soient $2 \leq j, l \leq J$, $j \neq l$ des entiers. Soit $b = (b_1, \dots, b_J)$ appartenant à \mathcal{A}_b , on note par (b_j^1, b_j^2) les coordonnées cartésiennes de b_j dans \mathbb{R}^2 . Pour tout $j = 2 \dots J$, les dérivées partielles d'ordre 1 de $M_{2,\epsilon}$ sont,

$$\left(\begin{array}{c} \frac{d}{db_j^1} \\ \frac{d}{db_j^2} \end{array} \right) M_{2,\epsilon}(\alpha_2) = -\frac{2\Re}{J} \int_{|\omega| \leq \omega_\epsilon} \frac{i}{\hat{a}_{j,\epsilon}} A_{-\hat{\theta}_{j,\epsilon}} \omega \hat{Z}_j(\omega) \overline{\left(\frac{1}{J} \sum_{l=1}^J \hat{Z}_l(\omega) \right)} d\omega,$$

et celles d'ordre 2 sont,

$$\begin{aligned} \left(\begin{array}{cc} \frac{d}{db_j^1 db_j^1} & \frac{d}{db_j^1 db_j^2} \\ \frac{d}{db_j^2 db_j^1} & \frac{d}{db_j^2 db_j^2} \end{array} \right) M_{2,\epsilon}(\alpha_2) &= \frac{2\Re}{J} \int_{|\omega| \leq \omega_\epsilon} \left(\frac{1}{\hat{a}_{j,\epsilon}} A_{-\hat{\theta}_{j,\epsilon}} \omega \right) \left(\frac{1}{\hat{a}_{j,\epsilon}} A_{-\hat{\theta}_{j,\epsilon}} \omega \right)^T \hat{Z}_j(\omega) \overline{\left(\frac{1}{J} \sum_{l=1}^J \hat{Z}_l(\omega) \right)} \\ &\quad - \frac{2\Re}{J} \int_{|\omega| \leq \omega_\epsilon} \left(\frac{1}{\hat{a}_{j,\epsilon}} A_{-\hat{\theta}_{j,\epsilon}} \omega \right) \left(\frac{1}{\hat{a}_{j,\epsilon}} A_{-\hat{\theta}_{j,\epsilon}} \omega \right)^T \left| \hat{Z}_j(\omega) \right|^2 d\omega, \\ \left(\begin{array}{cc} \frac{d}{db_j^1 db_l^1} & \frac{d}{db_j^1 db_l^2} \\ \frac{d}{db_j^2 db_l^1} & \frac{d}{db_j^2 db_l^2} \end{array} \right) M_{2,\epsilon}(\alpha_2) &= -\frac{2}{J} \int_{|\omega| \leq \omega_\epsilon} \left(\frac{1}{\hat{a}_{j,\epsilon}} A_{-\hat{\theta}_{j,\epsilon}} \omega \right) \left(\frac{1}{\hat{a}_{l,\epsilon}} A_{-\hat{\theta}_{l,\epsilon}} \omega \right)^T \hat{Z}_j(\omega) \overline{\hat{Z}_l(\omega)} \\ &\quad - \frac{2}{J} \int_{|\omega| \leq \omega_\epsilon} \left(\frac{1}{\hat{a}_{l,\epsilon}} A_{-\hat{\theta}_{l,\epsilon}} \omega \right) \left(\frac{1}{\hat{a}_{j,\epsilon}} A_{-\hat{\theta}_{j,\epsilon}} \omega \right)^T \hat{Z}_l(\omega) \overline{\hat{Z}_j(\omega)} d\omega. \end{aligned}$$

Proposition 4.3. Sous les hypothèses du théorème 4.4, la variable $\epsilon^{-1} \nabla M_{2,\epsilon}(\alpha_2^*)$ est uniformément tendue.

Preuve de la proposition 4.3. On reprend la décomposition du critère introduite dans la preuve du théorème 4.3. On étudie séparément la convergence des termes $\epsilon^{-1} \nabla D_{2,\epsilon}$, $\nabla L_{2,\epsilon}$, et $\epsilon \nabla Q_{2,\epsilon}$. Afin de simplifier les calculs par la suite, on utilise les notations suivantes :

$$\frac{d}{db_j} N_\epsilon(b) = \left(\begin{array}{c} \frac{d}{db_j^1} \\ \frac{d}{db_j^2} \end{array} \right) M_{2,\epsilon}(\alpha_2), \quad \forall b \in \mathcal{A}_b, \quad \forall j = 1, \dots, J.$$

D'après le lemme 4.2, nous avons pour $k = 2 \dots J$:

$$\begin{aligned} \frac{d}{db_j} D_{2,\epsilon}(\alpha_2^*) &= \frac{2\Re}{J^2} \sum_{l=1}^J \int_{|\omega| \leq \omega_\epsilon} \frac{i}{\hat{a}_{j,\epsilon}} A_{-\hat{\theta}_{j,\epsilon}} \omega \left[|\hat{f}(\omega)|^2 \right. \\ &\quad \left. - \left(\frac{a_j^* a_l^*}{\hat{a}_{j,\epsilon} \hat{a}_{l,\epsilon}} \right)^2 \hat{f} \left(\frac{a_j^*}{\hat{a}_{j,\epsilon}} A_{\theta_j^* - \hat{\theta}_{j,\epsilon}} \omega \right) \overline{\hat{f} \left(\frac{a_l^*}{\hat{a}_{l,\epsilon}} A_{\theta_l^* - \hat{\theta}_{l,\epsilon}} \omega \right)} \right] d\omega. \end{aligned}$$

Par la méthode Delta et par le fait que la fonction f soit 1-Lipschitz, on a :

$$\epsilon^{-1} \frac{d}{db_j} D_{2,\epsilon}(\alpha_2^*) = \mathcal{O}_{\mathbb{P}} \left(\int_{|\omega| \leq \omega_\epsilon} |\omega|^2 |\hat{f}(\omega)|^2 + |\omega| |\hat{f}(\omega)| d\omega \right).$$

Ainsi par les hypothèses (4.43) et (4.45), on déduit que $\epsilon^{-1}\nabla D_{2,\epsilon}(\alpha_2^*)$ est un vecteur uniformément tendu.

Concernant l'étude du gradient de $L_{2,\epsilon}$, les dérivées partielles sont égales pour $k = 2 \dots J$ à :

$$\frac{d}{db_j} L_{2,\epsilon}(b^*) = \frac{2}{J^2} \sum_{l=1}^J (V_{j,l} - \tilde{V}_{j,l}),$$

où

$$\begin{aligned} V_{j,l} &= \Re \int_{|\omega| \leq \omega_\epsilon} \frac{i}{\hat{a}_{j,\epsilon}} A_{-\hat{\theta}_{j,\epsilon}} \omega \left[\left(\frac{a_j^*}{\hat{a}_{j,\epsilon} \hat{a}_{l,\epsilon}} \right)^2 \hat{f} \left(\frac{a_j^*}{\hat{a}_{j,\epsilon}} A_{\theta_j^* - \hat{\theta}_{j,\epsilon}} \omega \right) \overline{\hat{W}_l \left(\frac{1}{\hat{a}_{l,\epsilon}} A_{-\hat{\theta}_{l,\epsilon}} \omega \right)} e^{-i\omega \cdot \tilde{\beta}_l^*} \right] d\omega. \\ \tilde{V}_{j,l} &= \Re \int_{|\omega| \leq \omega_\epsilon} \frac{i}{\hat{a}_{j,\epsilon}} A_{-\hat{\theta}_{j,\epsilon}} \omega \left[\left(\frac{a_l^*}{\hat{a}_{j,\epsilon} \hat{a}_{l,\epsilon}} \right)^2 \hat{f} \left(\frac{a_l^*}{\hat{a}_{l,\epsilon}} A_{\theta_l^* - \hat{\theta}_{l,\epsilon}} \omega \right) \overline{\hat{W}_j \left(\frac{1}{\hat{a}_{j,\epsilon}} A_{-\hat{\theta}_{j,\epsilon}} \omega \right)} e^{-i\omega \cdot \tilde{\beta}_j^*} \right] d\omega. \end{aligned}$$

Ainsi, il suffit de montrer que les variables $V_{j,l}$ converge en loi vers des variables gaussiennes et sont asymptotiquement centrées pour un j et un l fixés. Par la méthode Delta et en utilisant le fait que \hat{f} est au moins 1-Lipschitz, on déduit que :

$$\begin{aligned} V_{j,l} &= \int_{\mathcal{D}} \Re \int_{|\omega| \leq \omega_\epsilon} \frac{i}{a_j^*} A_{-\theta_j^*} \omega \left[\frac{1}{a_l^{*2}} \hat{f}(\omega) e^{i\omega \cdot \left(\frac{1}{a_l^*} A_{-\theta_l^*} x - \beta_l^* \right)} \right] d\omega dW_l(x) \\ &\quad + \mathcal{O}_{\mathbb{P}} \left(\int_{|\omega| \leq \omega_\epsilon} \epsilon |\omega|^2 |\hat{f}(\omega)| + \epsilon |\omega|^2 d\omega \right). \end{aligned}$$

Le second terme converge en probabilité vers zéro à cause des hypothèses (4.43) et (4.45). Alors que le premier terme converge vers un vecteur gaussien centré. Par conséquent, $\nabla L_{2,\epsilon}(\alpha_2^*)$ converge en loi vers un vecteur gaussien asymptotiquement centré.

Le dernier terme à étudier est $\epsilon \nabla Q_{2,\epsilon}(\alpha_2^*)$. On montre que cette variable converge en probabilité vers zéro. En effet, cette variable est la combinaison linéaire des variables suivantes

$$\epsilon \Re \int_{|\omega| \leq \omega_\epsilon} \frac{i}{\hat{a}_{j,\epsilon}} A_{-\hat{\theta}_{j,\epsilon}} \hat{W}_j \left(\frac{1}{\hat{a}_{j,\epsilon}} A_{-\hat{\theta}_{j,\epsilon}} \omega \right) e^{i\omega \cdot \tilde{\beta}_j^*} \overline{\hat{W}_l \left(\frac{1}{\hat{a}_{l,\epsilon}} A_{-\hat{\theta}_{l,\epsilon}} \omega \right)} e^{-i\omega \cdot \tilde{\beta}_l^*} d\omega,$$

avec $j = 1 \dots J$, $l = 1 \dots J$, $j \neq l$. Or pour un j et un l fixés, cette variable a le même comportement asymptotique que $\mathcal{O}_{\mathbb{P}}(\epsilon \omega_\epsilon^4)$. Donc par l'hypothèse (4.45), cette variable converge en probabilité vers zéro. \square

Proposition 4.4. Soit $\mathcal{A}_{2,\epsilon}^{loc}$ l'espace suivant :

$$\mathcal{A}_{2,\epsilon}^{loc} = \{ \alpha_2 \in \mathcal{A}_2, \|\alpha_2 - \alpha_2^*\| \leq \|\hat{\alpha}_{2,\epsilon} - \alpha_2^*\| \}.$$

Sous les hypothèses du théorème 4.4, il existe une matrice inversible H_b telle que

$$\sup_{(\bar{\alpha}_{2,\epsilon}) \in \mathcal{A}_{2,\epsilon}^{loc}} \|\nabla^2 M_{2,\epsilon}(\bar{\alpha}_{2,\epsilon}) - H_b\| = o_{\mathbb{P}}(1)$$

Preuve de la proposition 4.4. en utilisant les mêmes arguments que ceux de la proposition 4.2, les hypothèses du théorème 4.4 assurent l'uniforme convergence de la matrice hessienne de $M_{2,\epsilon}$ vers la matrice hessienne de M_2 . La matrice H_b est alors la valeur de la matrice hessienne de M_2 en α_2^* . Plus exactement cette matrice est égale à

$$H_b = \frac{2}{J} \begin{pmatrix} \frac{1}{a_2^{*2}} A_{-\theta_2^*} G A_{\theta_2^*} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{1}{a_J^{*2}} A_{-\theta_J^*} G A_{\theta_J^*} \end{pmatrix} - \frac{2}{J^2} \begin{pmatrix} \frac{1}{a_2^*} A_{-\theta_2^*} \\ \vdots \\ \frac{1}{a_J^{*2}} A_{-\theta_J^*} \end{pmatrix} \begin{pmatrix} \frac{1}{a_2^*} A_{-\theta_2^*} G \\ \vdots \\ \frac{1}{a_J^{*2}} A_{-\theta_J^*} G \end{pmatrix}^T,$$

où G est la matrice de Gramm des dérivées partielles de f :

$$G = \begin{pmatrix} \int_{\mathbb{R}^2} |\partial_1 f(x)|^2 \frac{dx}{2\pi} & \int_{\mathbb{R}^2} \partial_1 f(x) \partial_2 f(x) \frac{dx}{2\pi} \\ \int_{\mathbb{R}^2} \partial_1 f(x) \partial_2 f(x) \frac{dx}{2\pi} & \int_{\mathbb{R}^2} |\partial_2 f(x)|^2 \frac{dx}{2\pi} \end{pmatrix}.$$

Montrons que H_b est inversible. Comme les fonctions $\partial_1 f$ et $\partial_2 f$ sont linéairement indépendantes dans l'espace vectoriel $\mathbb{L}^2(\mathcal{D})$, la matrice G est inversible. Par conséquent, la matrice H_b est inversible, et sa matrice inverse est égale à

$$H_b^{-1} = \frac{J}{2} \begin{pmatrix} a_2^{*2} A_{-\theta_2^*} G^{-1} A_{\theta_2^*} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & a_J^{*2} A_{-\theta_J^*} G^{-1} A_{\theta_J^*} \end{pmatrix} + \frac{J}{2} \begin{pmatrix} a_2^* A_{-\theta_2^*} \\ \vdots \\ a_J^{*2} A_{-\theta_J^*} \end{pmatrix} \begin{pmatrix} a_2^* A_{-\theta_2^*} G^{-1} \\ \vdots \\ a_J^{*2} A_{-\theta_J^*} G^{-1} \end{pmatrix}^T.$$

□

Chapitre 5

Alignement de séries temporelles

Sommaire

5.1	Présentation du modèle	160
5.2	Estimation des décalages	161
5.3	Propositions et lemmes techniques	171

Résumé : On observe un nombre fixé de signaux discrets qui diffèrent entre eux par un décalage temporel. Puisque la base de Fourier diagonalise l'opérateur de translation, on propose une procédure statistique basée sur la transformée de Fourier du signal pour estimer les paramètres de décalages.

Mots clés : Méthode d'alignement, Estimation semi-paramétrique, Transformée de Fourier.

Introduction

Dans la théorie du signal, on distingue deux catégories de signaux. La première est l'ensemble des signaux discrets qui sont modélisés par des suites de réels $(f_l)_{l \in \mathbb{Z}}$. La seconde est l'ensemble des signaux continus qui sont modélisés par des fonctions à valeurs réelles $f : \mathbb{R} \rightarrow \mathbb{R}$. Souvent, les signaux observés ont approximativement la même forme, mais ne sont pas parfaitement alignés temporellement. Pour des signaux discrets, ceci peut être modélisé par la famille d'opérateurs suivante :

$$T_\theta : f = (f_l)_{l \in \mathbb{Z}} \in \mathcal{S}(\mathbb{Z}) \longrightarrow T_\theta(f) = (f_{l-\theta})_{l \in \mathbb{Z}} \in \mathcal{S}(\mathbb{Z}), \quad \theta \in \mathbb{Z},$$

où $\mathcal{S}(\mathbb{Z})$ est l'ensemble de tous les signaux discrets. Une illustration de ce problème peut être rencontrée dans l'étude des signaux cardiaques, des ondes ECG (cf. [Laciar et al. \(2003\)](#)) ou dans l'analyse de la voix (cf. [Sakoe et Chiba \(1978\)](#)). Il est bien connu qu'à moins d'enlever les variations en temps, la moyenne des signaux conduit à une perte d'information : elle ne capture pas la taille des sommets et sous-estime leurs pentes (voir [Kneip et al. \(2000\)](#) et [Ramsay et Li \(1998\)](#)). Un travail préliminaire est alors l'alignement des signaux et donc de l'estimation des retards ou des décalages θ entre les signaux.

Dans le traitement des signaux, beaucoup de méthodes existent. Nous passons en revue quelques unes de celles-ci. Pour aligner des électrocardiogrammes, la technique la plus employée est la corrélation croisée ou "cross correlation" (voir [Laciar et al. \(2003\)](#)). Cette technique consiste à maximiser la corrélation croisée entre deux signaux (le décalage θ entre ces signaux étant le maximum de la fonction de corrélation croisée). D'autres méthodes sont employées comme la méthode de double niveau ("the double level method"), les intégrales normalisées ou des méthodes de filtrage dont l'une est connue sous le nom de "the match filtering method". Une étude comparative de ces méthodes est donnée par [Jane et al. \(1991\)](#). Par ailleurs, une approche bayésienne pour l'estimation des décalages est développée par [Wahlberg et Salomonsson \(2003\)](#).

Dans l'analyse fonctionnelle, quelques méthodes ont été développées. Par exemple, en ingénierie, [Sakoe et Chiba \(1978\)](#) développent une approche, appelée "dynamic time warping", pour aligner deux signaux avec des dynamiques différentes. Le principe de cette méthode est de minimiser une fonction de coût sur un ensemble de chemins de déformation (cf. [Wang et Gasser \(1997\)](#)). Cette méthode a été principalement appliquée à l'analyse de la voix et à la reconnaissance auditive des mots. Une autre approche est d'appliquer des concepts tel que l'analyse par composante principale ([Rao \(1958\)](#) et [Silverman \(1995\)](#)). D'autre part, [Kneip et Gasser \(1992, 1995\)](#) ont développé une technique, nommée "structural analysis", qui aligne les signaux (continus) en utilisant les extrema ou marqueurs.

Pour les signaux continus, une approche semi-paramétrique est proposée par [Lawton et al. \(1972\)](#). Les auteurs définissent une méthode d'estimation, SEMOR, pour les décalages en temps et pour l'estimation de la forme générale du signal. La consistance et la convergence de cette méthode a été prouvée par [Kneip et Gasser \(1988\)](#). [Härdle et Marron \(1990\)](#) définissent une méthode d'estimation utilisant un estimateur à noyau à la place de l'approximation polynomiale de la méthode SEMOR : l'estimateur des décalages ainsi construit est \sqrt{n} -consistant. Plus récemment, [Gamboa et al. \(2005\)](#) construisent un estimateur lui aussi \sqrt{n} -consistant pour les décalages en temps en utilisant la transformée de Fourier discrète. Le chapitre 2 généralise cette méthode pour des modèles plus complets. On a vu que si le processus de contraste est correctement défini, l'estimateur construit était asymptotiquement efficace.

Ce chapitre a pour but de développer la théorie duale de [Gamboa et al. \(2005\)](#) et du chapitre 2 pour les séries temporelles. Ce ne sont plus des décalages en temps pour des signaux continus que l'on cherche à estimer, mais pour des signaux à temps discrets. Notre procédure d'estimation est donc basée sur la M -estimation et la transformée de Fourier des données. On s'est intéressé ici aux signaux ergodiques qui admettent une distribution spectrale. Ces distributions spectrales sont définies à partir de la convergence étroite de l'autopériodogramme (la norme au carré de la transformée de Fourier des signaux). Puisque la base de Fourier diagonalise les opérateurs $(T_\theta)_{\theta \in \mathbb{Z}}$, deux signaux, en décalages l'un de l'autre, admettent la même distribution spectrale. Par contre, les transformées de Fourier de ces signaux diffèrent asymptotiquement d'un nombre complexe de module 1 dont l'argument dépend du décalage à estimer. Une heuristique de ce problème est présentée par [Bloomfield \(2000\)](#). Notre processus de contraste pour estimer les décalages en temps entre les signaux utilise les deux aspects que l'ont vient de développer : la convergence étroite des autopériodogrammes et la propriété des transformées de Fourier sur les opérateurs de translation. On a constaté par la suite que ce processus de contraste est en fait proche de la méthode de "cross correlation" ([Laciar et al. \(2003\)](#)).

Le modèle étudié est un modèle de régression :

$$Y_{j,l} = f_{l-\theta_j^*} + \sigma_l \epsilon_{j,l}, \quad |l| \leq n, \quad j = 1 \dots J.$$

On cherche à aligner les signaux discrets $(f_{l-\theta_j^*})_{l \in \mathbb{Z}}$, c'est-à-dire à estimer les paramètres $\theta^* = (\theta_j^*)_{j=1, \dots, J}$ à partir des données $Y_{j,l}$. Ces signaux sont supposés admettre une mesure spectrale. Les termes de bruits $(\epsilon_{j,l})$ sont des variables centrées, de variance égale à 1. Afin que la procédure d'estimation soit consistante, on est conduit à supposer que la suite de réels positifs $(\sigma_l)_l$ converge vers zéro en moyenne de Cesaro. D'autre part, il est remarquable que le processus de contraste utilisé pour l'estimation de θ^*

est défini sur un compact de \mathbb{R}^J et pas seulement sur un sous-ensemble de \mathbb{Z}^J . Ainsi les outils d'optimisation différentielle peuvent être utilisés.

Ce chapitre s'articule de la manière suivante. La section 5.1 décrit le modèle étudié et les outils mathématiques utilisés. Dans la section 5.2, on étudie la consistance de l'estimateur des décalages (le théorème 5.1), ainsi que sa vitesse de convergence (le théorème 5.2). Dans cette même section, on présente des applications de cette méthode sur des données simulées.

5.1 Présentation du modèle

Description générale du modèle.

Les données $(Y_{j,l})_{1 \leq j \leq J, |l| \leq n}$ sont les observations bruitées de J signaux : j est l'indice spécifiant le signal, l est l'indice des temps. Ces signaux sont supposés identiques à un décalage en temps près. Le modèle étudié est le modèle de régression suivant :

$$Y_{j,l} = f_{l-\theta_j^*} + \sigma_l \epsilon_{j,l}, \quad |l| \leq n, \quad j = 1, \dots, J. \quad (5.1)$$

Le signal $(f_l)_{l \in \mathbb{Z}}$ est le signal d'intérêt. Il vérifie une propriété d'ergodicité (5.3) définie par la suite. Les termes $(\epsilon_{j,l})_{j,l}$ sont des variables centrées de variance 1 indépendantes et identiquement distribuées. La suite de réels positifs (σ_l^2) est telle qu'elle converge en moyenne de Cesaro vers 0 :

$$\frac{1}{n} \sum_{|l| \leq n} \sigma_l^2 = o(1). \quad (5.2)$$

L'objectif est donc d'estimer les entiers relatifs $(\theta_j^*)_{j=1, \dots, J}$.

Contraintes d'identifiabilité.

Au vu des données il est raisonnable de supposer que les paramètres de décalages (θ_j^*) soient dans un sous-ensemble compact de \mathbb{R}^J :

$$\theta_{\min} \leq \theta_j^* \leq \theta_{\max}, \quad j = 1, \dots, J,$$

où θ_{\min} et θ_{\max} sont des constantes fixées par l'utilisateur. On suppose sans perte de généralités que $\theta_{\min} = -\theta_{\max}$. Or le modèle (5.1) tel qu'il est présenté n'est pas identifiable. En effet, soit $v \in \mathbb{Z}^*$. Si on considère le signal $(g_l = f_{l-v})$ à la place du signal (f_l) , on a le même modèle sauf que le vecteurs des décalages $(\theta_j^*)_{j=1, \dots, J}$ est remplacé par $(\theta_j^* + v)_{j=1, \dots, J}$. On est donc amené à réduire l'espace des paramètres Θ en posant des contraintes d'identifiabilité. Cet espace doit vérifier les conditions suivantes

- Θ est un compact (afin d'assurer la consistance)
- le vrai paramètre $\theta^* = (\theta_j^*)_{j=1,\dots,J}$ appartient à Θ
- pour tout $v \in \mathbb{Z}^*$, on ne peut pas définir un modèle équivalent avec le paramètre $(\theta_j^* + v)_{j=1,\dots,J} \in \Theta$ à la place de $\theta^* = (\theta_j^*)_{j=1,\dots,J}$.

Dans ce chapitre, on pose la contrainte d'identifiabilité $\theta_1^* = 0$. L'espace des paramètres Θ est alors le suivant

$$\Theta = \{\theta \in [-\theta_{\max}, \theta_{\max}]^J, \text{ tel que } \theta_1 = 0\}.$$

La contrainte ($\theta_1 = 0$) signifie que le premier signal est pris pour référence pour calculer les décalages.

Distribution spectrale.

Comme il a été mentionné dans l'introduction, le processus de contraste pour l'estimation du paramètre θ^* est défini à partir de la transformée de Fourier des données. Soit J_n la transformée de Fourier du signal discret (f_l) :

$$J_n(\lambda) = \frac{1}{\sqrt{n}} \sum_{|l| \leq n} f_l e^{-il\lambda}, \quad \lambda \in [-\pi, \pi].$$

L'autoperiogramme associé au signal (f_l) est alors la fonction I_n suivante :

$$I_n(\lambda) = \frac{1}{2\pi n} \left| \sum_{|l| \leq n} f_l e^{-il\lambda} \right|^2 = \frac{1}{2\pi} |J_n(\lambda)|^2, \quad \lambda \in [-\pi, \pi].$$

On suppose que le signal (f_l) vérifie la propriété d'ergodicité suivante : il existe une mesure F sur $[-\pi, \pi]$ telle que pour toute fonction bornée à valeur réelle g sur $[-\pi, \pi]$, on a :

$$\lim_{n \rightarrow +\infty} \int_{-\pi}^{\pi} g(\lambda) I_n(\lambda) d\lambda = \int_{-\pi}^{\pi} g(\lambda) dF(\lambda). \quad (5.3)$$

La fonction F est distribution spectrale associée au signal (f_l) . La propriété 5.3 est vérifiée par exemple pour des processus (f_l) stationnaires, voir par exemple [Bercu et Prieur \(2006\)](#) ou [Lopes et Lopes \(2002\)](#).

5.2 Estimation des décalages

5.2.1 Estimation par processus de contraste

Définition du processus de contraste.

L'estimateur des décalages θ^* est défini comme le maximum d'un processus de contraste M_n . La définition de ce processus est largement inspirée du chapitre 2 où

l'on considérerait des signaux continus. Il est défini à partir de la transformée de Fourier des données $(Y_{j,l})$:

$$d_j(\lambda) = \frac{1}{\sqrt{n}} \sum_{|l| \leq n} Y_{j,l} e^{-\lambda l}, \quad \lambda \in [-\pi, \pi], \quad j = 1, \dots, J. \quad (5.4)$$

Comme le fait remarquer [Bloomfield \(2000\)](#), si le signal $(Y_{j,l})_l$ est décalé en temps de θ_j^* unités par rapport au signal $(Y_{1,l})_l$, alors asymptotiquement les transformées de Fourier d_j et d_1 sont reliés par l'équation suivante :

$$d_j(\lambda) \cong d_1(\lambda) e^{-i\lambda\theta_j^*}. \quad (5.5)$$

L'égalité n'est pas exacte car les derniers termes de la somme (5.4) ne sont pas identiques d'une part, et que les signaux sont bruités. Ainsi en rephasant les transformées de Fourier, le processus de contraste considéré est :

$$M_n(\theta) = \int_{-\pi}^{\pi} \left| \frac{1}{J} \sum_{j=1}^J d_j(\lambda) e^{i\lambda\theta_j} \right|^2 \frac{d\lambda}{2\pi}, \quad \theta \in \Theta. \quad (5.6)$$

On remarque que ce processus n'est pas défini seulement pour les vecteurs d'entiers relatifs, mais aussi pour des réels.

D'autre part, comme pour le chapitre 2, on peut généraliser la définition du processus et étendre les théorèmes qui suivent pour le modèle où les signaux diffèrent non seulement en phase mais en amplitude :

$$Y_{j,l} = a_j^* f_{l-\theta_j^*} + \epsilon_{j,l}, \quad |l| \leq n, \quad j = 1, \dots, J,$$

avec (a_j^*) sont des réels positifs. Ce dernier modèle est identifiable si l'on suppose que $\sum_{j=1}^J a_j^{*2} = J$ par exemple.

Consistance de l'estimateur

On étudie maintenant les propriétés asymptotiques de l'estimateur minimisant le processus de contraste 5.6.

Théorème 5.1 (Consistance). *Soit $\hat{\theta}_n$ l'estimateur défini comme :*

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta} M_n(\theta).$$

Le modèle est tel que la suite (σ_l) converge en moyenne de Cesaro 5.2. De plus si l'autoperiogramme I_n converge étroitement vers la mesure spectrale F , dont l'intérieur du support est supposé non vide, alors $\hat{\theta}_n$ converge en probabilité vers θ^ .*

Démonstration. La preuve suit les grandes lignes de la convergences des M -estimateurs (voir le chapitre 1 pour plus de détails). Il suffit de montrer que :

i) M_n converge uniformément en probabilité vers une fonction de contraste M (lemme 5.1). Par l'hypothèse d'ergodicité (5.3) et par la propriété de la transformée de Fourier (5.5), cette fonction M est,

$$M(\theta) = \int_{-\pi}^{\pi} \left| \frac{1}{J} \sum_{j=1}^J e^{i\lambda(\theta_j - \theta_j^*)} \right|^2 dF(\lambda).$$

ii) La fonction de contraste M possède un unique maximum (lemme 5.2). □

Dans le modèle de régression étudié, l'hypothèse (5.2) est nécessaire pour assurer la convergence du processus de contraste. En effet, il faut que l'autopériodogramme $|d_1(\lambda)|^2$ converge étroitement vers F . Dans ce cas, la variable aléatoire suivante doit converger en probabilité vers zéro :

$$\frac{1}{n} \sum_{|l| \leq n} \sigma_l^2 \epsilon_{j,l}^2.$$

Ce qui revient à supposer (5.2).

Plus précisément, cette convergence en moyenne de Cesaro (5.2) signifie que la suite $(\sigma_l^2)_l$, privé d'un ensemble négligeable de terme, converge vers zéro. Même si cette hypothèse n'est pas aussi restrictive que la convergence simple de $(\sigma_l^2)_l$ vers zéro, elle limite la robustesse du critère. Il serait sans doute préférable de pouvoir éviter à poser une telle hypothèse. Une solution pourrait être l'emploi d'une fenêtre dans la définition de l'autopériodogramme.

En effet, il est bien connu que l'autopériodogramme d'un signal est généralement irrégulier en dépit du fait que la mesure F admette une densité spectrale régulière : l'autopériodogramme n'est pas un bon estimateur de cette densité spectrale. Pour régulariser cet estimateur, il est courant d'introduire une fenêtre $(h_{l,n})_l$, et de considérer l'estimateur de la densité spectrale suivant :

$$\left| \frac{1}{\sqrt{n}} \sum_{|l| \leq n} h_{l,n} f_l \right|^2.$$

Alors la condition que $\sum_{|l| \leq n} h_{l,n}^2/n$ converge vers zéro serait suffisante pour contrôler le bruit. Le problème est alors de trouver une telle suite de poids qui n'interférera pas dans la limite du critère.

Concernant l'implémentation du critère, le calcul du processus de contraste utilise les résultats de Lopes et Lopes (2002) ou ceux exposés dans Brillinger (2001). Par

contre, la sous-section suivante met en évidence que dans la pratique l'estimateur qui maximise la fonction (5.6) est difficile à déterminer : la fonction M_n a bien un unique maximum global, mais elle a aussi de multiples maximum locaux. Ainsi les méthodes d'optimisation différentielle ne sont guère utiles, si on ne connaît pas un estimateur préliminaire ou si on ne trace pas graphiquement le processus de contraste M_n (ce qui limite considérablement le nombre de signaux à recalculer). Pour contourner ce problème, on peut penser à localiser le processus de contraste avec une fonction de poids $\omega : [-\pi, \pi] \rightarrow \mathbb{R}$ continue :

$$M_{n,\omega}(\theta) = \int_{-\pi}^{\pi} \left| \frac{1}{J} \sum_{j=1}^J d_j(\lambda) e^{i\lambda\theta_j} \right|^2 \omega(\lambda) \frac{d\lambda}{2\pi}, \quad \theta \in \Theta. \quad (5.7)$$

Il reste à savoir quelle fonction de poids choisir.

Le théorème qui suit étudie la vitesse de convergence de l'estimateur des décalages. Cette vitesse ne peut qu'être plus petite que $n^{1/2}$, car dans le cas contraire le signal devrait être de carré sommable. De plus, cette vitesse dépend du contrôle de la suite des variances (σ_l^2) du bruit.

Théorème 5.2 (Vitesse de convergence). *Soit $\alpha \in]0, 1[$ tel que la suite des variances vérifie :*

$$\frac{1}{n^\alpha} \sum_{i=1}^n \sigma_i^2 = \mathcal{O}(1). \quad (5.8)$$

Supposons que le signal f_l est borné et que les hypothèses du théorème 5.1 soient vérifiées. Alors, les fluctuations $n^{(1-\alpha)/2}(\hat{\theta}_n - \theta^)$ sont uniformément tendues :*

$$n^{(1-\alpha)/2}(\hat{\theta}_n - \theta^*) = \mathcal{O}_{\mathbb{P}}(1).$$

Démonstration. Comme $\hat{\theta}_n$ est le maximum du processus de contraste M_n , on a :

$$\nabla M_n(\hat{\theta}_n) = 0.$$

Par un développement de Taylor du gradient de M_n , il existe $\bar{\theta}_n$ dans un voisinage de θ^* tel que :

$$\nabla^2 M_n(\bar{\theta}_n) n^{(1-\alpha)/2}(\hat{\theta}_n - \theta^*) = -\sqrt{n} \nabla M_n(\theta^*).$$

Par le lemme 5.4, le terme $\nabla^2 M_n(\bar{\theta}_n)$ converge en probabilité vers une matrice inversible. D'autre part, le lemme 5.3 établit que $n^{(1-\alpha)/2} \nabla M_n(\theta^*)$ est uniformément tendue. Donc, $n^{(1-\alpha)/2}(\hat{\theta}_n - \theta^*)$ est uniformément tendue. \square

Pour conclure cette partie, il est remarquable que le critère présenté se rapproche en fait du critère de corrélation croisée (Laciar et al., 2003). En effet pour $J = 2$ et

$\theta \in \mathbb{Z}$, le processus de contraste est égale à :

$$M_n(\theta) = \frac{1}{4n} \sum_{|l| \leq n} (|Y_{1,l}|^2 + |Y_{2,l}|^2) + \sum_{l=\max(-n, -n-\theta)}^{\min(n, n-\theta)} Y_{1,l} Y_{2,(l-\theta)}.$$

Le dernier terme de la somme est identique au numérateur de la corrélation croisée. Par conséquent, la valeur entière de l'estimateur de θ_2^* qui maximise le processus de contraste M_n est plus ou moins un l'entier qui maximise la corrélation croisée. Le processus de contraste développé ici généralise donc cette méthode pour un nombre fixé J de signaux. Bien entendu, pour que cette méthode soit praticable, le nombre de signaux J considérés ne doit pas être trop grands : l'espace d'optimisation croît de manière exponentielle avec J . Le corollaire 5.1 est une conséquence du théorème 5.1. Pour $J = 2$, il établit la consistance de l'estimateur défini comme le minimum de la corrélation croisée.

Corollaire 5.1. *Soit $\tilde{\theta}_n$ l'estimateur suivant,*

$$\tilde{\theta}_n = \arg \min_{\theta \in \Theta \cap \mathbb{Z}^J} M_n(\theta).$$

Sous les hypothèses du théorème 5.1, la suite d'entiers $\tilde{\theta}_n$ converge en probabilité vers θ^ . En particulier, puisque $\tilde{\theta}_n$ est à valeur dans un sous-ensemble de \mathbb{Z}^J , il existe un entier aléatoire \hat{N} tel que pour tout $n \geq \hat{N}$ $\tilde{\theta}_n = \theta^*$.*

5.2.2 Simulations

On présente maintenant des application de cette méthode sur des données simulées. Les signaux considérés sont successivement des signaux gaussiens indépendamment distribués, $MA(1)$, $MA(3)$, $AR(1)$ et sur un exemple de signal dynamique.

Nous allons développer les remarques faites dans la sous-section précédente, c'est-à-dire la multiplicité des maximum locaux, l'incidence du bruit, le choix d'une fonction de poids et la généralisation de la technique de corrélation croisée.

Application 1 : signal indépendant identiquement distribué

Dans ce premier exemple, on considère le signal f_l qui est la réalisation d'une suite de variables indépendantes gaussiennes centrées de variance 1. La distribution spectrale est alors la mesure de Lebesgue sur $[-\pi, \pi]$, $dF(\lambda) = d\lambda$. Le modèle étudié est le suivant :

$$\begin{cases} Y_{1,l} = f_l + \sigma \epsilon_{1,l} & l = -n \dots n \\ Y_{2,l} = f_{l-\theta_2^*} + \sigma \epsilon_{2,l} & l = -n \dots n \\ Y_{3,l} = f_{l-\theta_3^*} + \sigma \epsilon_{3,l} & l = -n \dots n \end{cases}$$

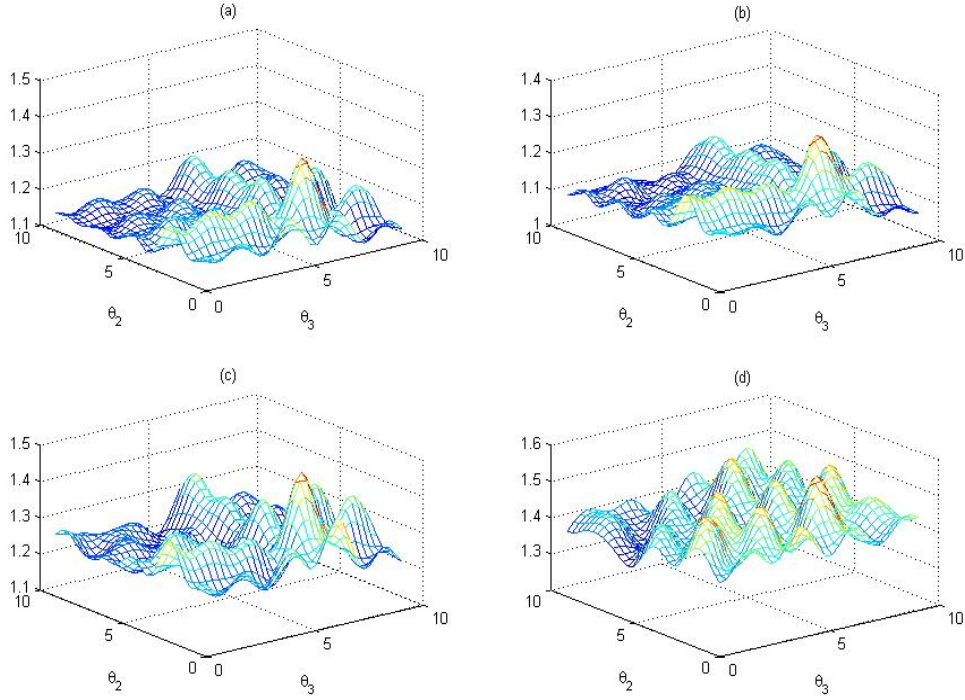


FIG. 5.1: Graphique du processus de contraste (5.6) pour des variables indépendantes identiquement distribuées ($J = 3$ et $n = 62$) et pour différentes valeurs de σ : (a) $\sigma = 0$, (b) $\sigma = 0.2$, (c) $\sigma = 0.4$, (d) $\sigma = 0.6$.

où $\theta^* = (0 \ 2 \ 6)$ et $\sigma \in [0, 1[$ est une constante positive fixée. La figure 5.1 présente les graphiques du processus de contraste pour $n = 62$ et différentes valeurs de σ .

La figure 5.1(a) illustre le fait que le processus de contraste possède un maximum global proche du point $(2, 6)$, et une multitude de maximum locaux répartis sur un réseau proche de $(2, 0)\mathbb{Z} \oplus (0, 6)\mathbb{Z}$. Ceci s'explique par la fonction de contraste qui est l'intégrale sur $[-\pi, \pi]$ de :

$$(\lambda, \theta_2, \theta_3) \rightarrow \frac{1}{9} \left| 1 + e^{i\lambda(\theta_2-2)} + e^{i\lambda(\theta_3-6)} \right|^2.$$

D'autre part, on remarque que si un bruit d'écart-type égal à 0.4 (voir la figure 5.1 (c)) n'est pas trop gênant pour déterminer le maximum global, la situation devient plus difficile dès qu'il excède 0.6 (voir la figure 5.1(d)). Le maximum global est toujours proche de $(2, 6)$, mais il se démarque nettement moins des maximum locaux. Pour illustrer ce fait, la figure 5.2 représente les diagrammes de Tukey pour les quatre modèles considérés dans la figure 5.1.

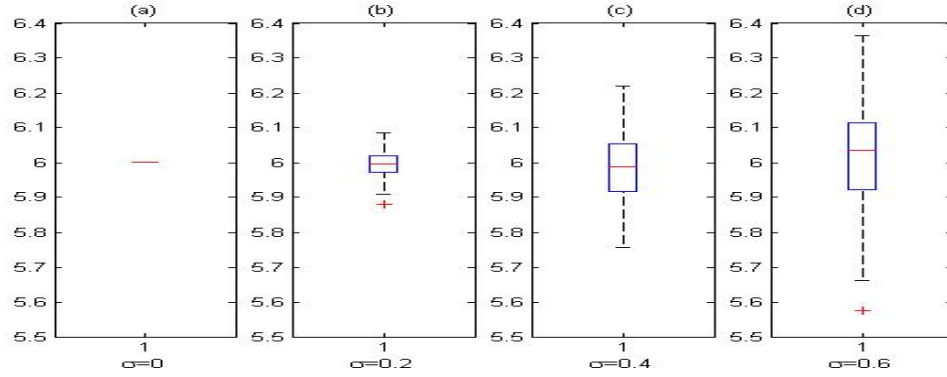


FIG. 5.2: Diagrammes de Tukey pour les estimateurs de $\theta_3^* = 5$ (définis à partir du processus (5.6)) construits à partir de 100 jeux de données dans le modèle où les signaux sont indépendants et identiquement distribués

Application 2 : un signal de moyenne mobile $MA(1)$ et $MA(3)$ gaussiens

Rappelons l'expression d'un processus $MA(3)$ gaussien centrée est :

$$f_l = z_l + \beta_1 z_{l-1} + \beta_2 z_{l-2} + \beta_3 z_{l-3},$$

où β_1 , β_2 et β_3 sont des réels fixés et (z_l) est une suite de variables indépendantes identiquement distribuées, $z_l \sim \mathcal{N}(0, 1)$. La distribution spectrale est alors $dF(\lambda) = |z_l + \beta_1 e^{i\lambda} + \beta_2 e^{i2\lambda} + \beta_3 e^{i3\lambda}|^2 d\lambda$. Le modèle étudié est le suivant :

$$\begin{cases} Y_{i,1} = f_l + 0.4\epsilon_{1,l} & l = -n \dots n \\ Y_{i,2} = f_{l-\theta_2^*} + 0.4\epsilon_{2,l} & l = -n \dots n \\ Y_{i,3} = f_{l-\theta_3^*} + 0.4\epsilon_{3,l} & l = -n \dots n \end{cases}$$

où $\theta^* = (0 \ 2 \ 6)$ et $n = 62$.

Le figure 5.3 représente le graphique du processus de contraste pour une $MA(1)$ de paramètre $(\beta_1, \beta_2, \beta_3) = (\sqrt{5}, 0, 0)$ et pour une $MA(3)$ de paramètre $(\beta_1, \beta_2, \beta_3) = (2, 0, -1)$. On observe que ces graphiques diffèrent de ceux présentés précédemment pour des signaux indépendants identiquement distribués. D'autre part, au sein d'une même famille, on constate une variabilité selon les paramètres (β_i) . Ceci reflète que la fonction de contraste dépend de la mesure spectrale. Ainsi, si l'on souhaite insérer une fonction de poids ω dans le contraste (voir le critère 5.7), il semble nécessaire de connaître non seulement la famille du signal (c'est-à-dire la forme générale de la densité spectrale), mais aussi avoir une idée des paramètres qui la régit.

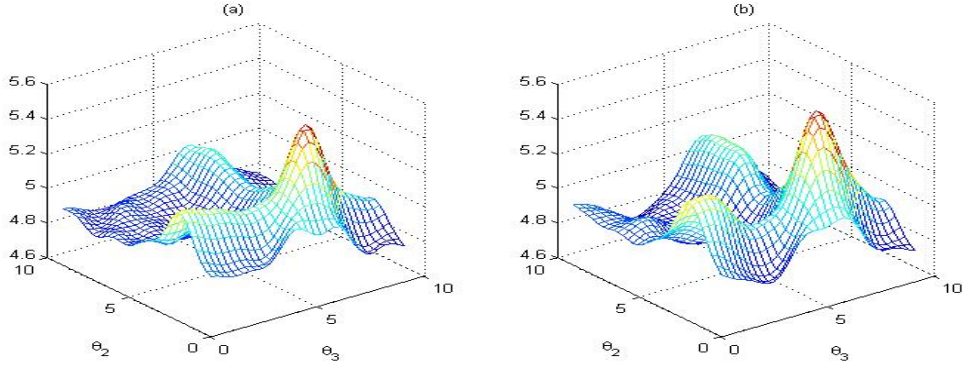


FIG. 5.3: Graphique du processus de contraste (5.6) pour des signaux à moyenne mobile ($J = 3$ et $n = 62$) : (a) $MA(1)$ avec $(\beta_1, \beta_2, \beta_3) = (\sqrt{5}, 0, 0)$, (b) $MA(3)$ avec $(\beta_1, \beta_2, \beta_3) = (2, 0, -1)$.

Application 3 : un signal autorégressif $AR(1)$ gaussien

Rappelons l'expression d'un processus $AR(1)$ gaussien centrée est :

$$f_l = z_l + \gamma f_{l-1},$$

où $\gamma \in (0, 1)$ est un réel fixé et (z_l) est une suite de variables indépendantes identiquement distribuées, $z_l \sim \mathcal{N}(0, 1)$. La distribution spectrale est alors $dF(\lambda) = d\lambda / |1 - \gamma e^{-\lambda}|^2$. Le modèle étudié est le suivant :

$$\begin{cases} Y_{i,1} = f_l + 0.1\epsilon_{1,l} & l = -n \dots n \\ Y_{i,2} = f_{l-\theta_2^*} + 0.1\epsilon_{2,l} & l = -n \dots n \\ Y_{i,3} = f_{l-\theta_3^*} + 0.1\epsilon_{3,l} & l = -n \dots n \end{cases}$$

où $\theta^* = (0 \ 2 \ 6)$ et $n = 62$.

La figure 5.4 illustre les limites de l'usage d'une fonction de poids ω dans le processus de contraste (5.7). Les figures 5.4(a), 5.4(c) et 5.4(d) représentent les processus de contraste pour des signaux $AR(1)$ avec $\gamma = 0.3$ pour différentes fonctions de poids ω . La figure 5.4(a) est le graphique du processus associé à la fonction de poids $\omega \equiv 1$: comme précédemment on observe un maximum global $(0 \ 2.0535 \ 6.0297)$ proche de θ^* et une multitude de maximum locaux. Les figures 5.4(c) et 5.4(d) représentent les graphiques des processus de poids associés à la fonction de poids,

$$\omega(\lambda) = \exp(-\lambda^2 / (2\delta)), \quad \lambda \in \mathbb{R},$$

avec $\delta = 2$ et $\delta = 1$ respectivement. Pour $\delta = 2$, la fonction de poids localise bien le processus : on observe un unique maximum global $(0 \ 1.9653 \ 5.9367)$ proche de θ^* .

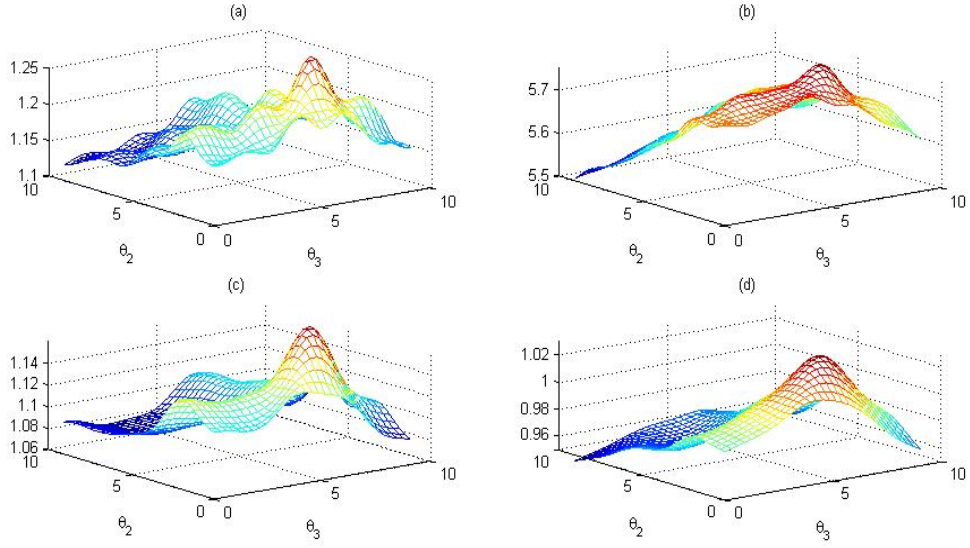


FIG. 5.4: Graphique du processus de contraste (5.7) pour des signaux $AR(1)$ ($J = 3$ et $n = 62$) : (a) $\gamma = 0.3$ avec $w \equiv 1$, (b) $\gamma = 0.7$ avec $w \equiv 1$, $\gamma = 0.3$ avec $w(\lambda) = e^{-\lambda^2/4}$, $\gamma = 0.3$ avec $w(\lambda) = e^{-\lambda^2/2}$.

Par contre, pour une fonction de poids plus concentrée au voisinage de l'origine ($\delta = 1$), la localisation est certainement meilleure mais le maximum du processus de contraste $(0 \ 1.8953 \ 5.7643)$ est plus éloigné de θ^* : un biais est introduit dans l'estimation des paramètres. Pour chacun des poids considérés dans les figures 5.4(a), 5.4(d) et 5.4(d), on a simulé 100 jeux de données et calculé pour chaque jeu l'estimateur de θ^* . La figure 5.5 représente les diagrammes de Tukey pour les estimateurs de $\theta_3^* = 5$ définis à partir des processus de contraste considérés pour les figures 5.4(a), 5.4(d) et 5.4(d) : plus, on concentre le poids, plus on perd de l'information sur le paramètre θ^* .

Enfin, comme on l'a remarqué dans l'exemple précédent, il peut exister une variabilité sur la forme des critères limites pour des signaux qui diffèrent seulement au niveau d'un paramètre : les figures 5.4(a) et 5.4(b) représentent respectivement les graphiques des processus de contraste (5.6) pour des $AR(1)$ de paramètres $\gamma = 0.3$ et $\gamma = 0.7$. Ainsi même si une fonction de poids conviendrait pour localiser le processus pour un certain signal, il se peut qu'elle ne soit pas optimale pour un autre signal de même nature avec un autre paramètre.

Application 4 : un exemple de signal dynamique

Le dernier exemple a pour but d'illustrer la généralisation de la méthode de corrélation croisée (corollaire 5.1). Les jeux de données sont générés à partir du modèle

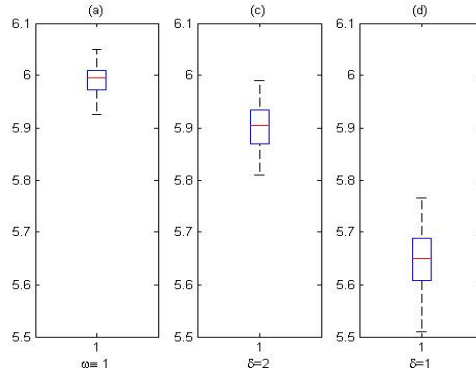


FIG. 5.5: Diagrammes de Tukey de l'estimateur de $\theta_3^* = 5$ (définis à partir du processus (5.7)) construits à partir de 100 jeux de données dans le modèle où les signaux sont des $AR(1)$ ($J = 3$ et $n = 62$) : (a) $\gamma = 0.3$ avec $w \equiv 1$, $\gamma = 0.3$ avec $w(\lambda) = e^{-\lambda^2/4}$, $\gamma = 0.3$ avec $w(\lambda) = e^{-\lambda^2/2}$.

suivant :

$$Y_{j,l} = f_{l-\theta_j^*} + 0.4\epsilon_{j,l}, \quad l = -n, \dots, n, \quad j = 1, \dots, 4,$$

où $\theta^* = (0, 2, 6, -1)$ et (f_l) est signal dynamique généré par la relation de récurrence suivante,

$$\begin{cases} f_{l+1} = 4f_l(1 - f_l) & \text{pour } l > -n \\ f_{-n} \sim \beta(1/2, 1/2), \end{cases}$$

où $\beta(1/2, 1/2)$ désigne la loi bêta de paramètre $(1/2, 1/2)$ et dont la densité $x \rightarrow (\pi x(1-x))^{-1/2}$, est à support dans $[0, 1]$. Pour différentes valeurs de n nous avons calculer $\tilde{\theta}_n$,

$$\tilde{\theta}_n = \arg \max \{M_n(\theta), \theta \in \{0\} \times \{-15, \dots, 15\}^3\}.$$

D'après le corollaire 5.1, il existe un entier aléatoire \hat{N} tel que $n > \hat{N}$, on avait que $\tilde{\theta}_n = \theta^*$. Sur 15 simulations effectués avec **Matlab**, on a observé que,

$$31 \leq \hat{N} \leq 45.$$

5.3 Propositions et lemmes techniques

Lemmes techniques pour le théorème 5.1

Remarque 5.1. Dans les preuves, on utilise les notations ci dessous. Soit $\theta \in \Theta$, on a pour $j \in \{1, \dots, J\}$ fixé :

$$\frac{1}{\sqrt{n}} \sum_{|l| \leq n} f_{l-\theta_j} e^{-i\lambda(l-\theta_j)} = \frac{1}{\sqrt{n}} \sum_{|l| \leq n} f_l e^{-i\lambda l} + R_n(\theta_j),$$

$$\text{où} \quad R_n(\theta_j) = \begin{cases} \frac{-1}{\sqrt{n}} \sum_{l=n-\theta_j+1}^n f_l e^{-i\lambda l} + \frac{1}{\sqrt{n}} \sum_{l=-n-\theta_j+1}^{-n+1} f_l e^{-i\lambda l} & \text{si } \theta_j > 0 \\ 0 & \text{si } \theta_j = 0 \\ \frac{-1}{\sqrt{n}} \sum_{l=-n}^{-n-\theta_j-1} f_l e^{-i\lambda l} + \frac{1}{\sqrt{n}} \sum_{l=n+1}^{n-\theta_j} f_l e^{-i\lambda l} & \text{si } \theta_j < 0 \end{cases}$$

Lemme 5.1 (La convergence uniforme en probabilité). Sous les hypothèses du théorème 5.1, on a :

$$\sup_{\theta \in \Theta} |M_n(\theta) - M(\theta)| = o_{\mathbb{P}}(1),$$

où

$$M(\theta) = \int_{-\pi}^{\pi} |\phi(\lambda, \theta)|^2 dF(\lambda) \quad \text{et} \quad \phi(\lambda, \theta) = \frac{1}{J} \sum_{j=1}^J e^{i\lambda(\theta_j - \theta_j^*)}.$$

Démonstration. On récrit le processus de contraste comme la somme de trois termes,

$$M_n(\alpha) = D_n(\alpha) + L_n(\alpha) + Q_n(\alpha),$$

où

- $D_n(\alpha)$ est la partie qui ne dépend que du signal (f_l) . En utilisant les notations introduites dans la remarque 5.1, D_n est égale à,

$$D_n(\theta) = \int_{-\pi}^{\pi} \left| \frac{1}{J\sqrt{n}} \sum_{j=1}^J \sum_{|l| \leq n} f_{l-\theta_j^*} e^{-i\lambda(l-\theta_j^*)} \right|^2 \frac{d\lambda}{2\pi}$$

$$= \int_{-\pi}^{\pi} |\phi(\lambda, \theta)|^2 I_n(\lambda) d\lambda \tag{5.9}$$

$$+ \int_{-\pi}^{\pi} \Re \left\{ \phi(\lambda, \theta) \sum_l J_n(\lambda) \frac{1}{J} \sum_{j=1}^J e^{i\lambda(\theta_j - \theta_j^*)} R_n(\theta_j^*) \right\} \frac{d\lambda}{2\pi} \tag{5.10}$$

$$+ \int_{-\pi}^{\pi} \left| \frac{1}{J} \sum_{j=1}^J e^{i\lambda(\theta_j - \theta_j^*)} R_n(\theta_j^*) \right|^2 \frac{d\lambda}{2\pi}. \tag{5.11}$$

- L_n est la partie qui dépend linéairement du bruit,

$$L_n(\theta) = 2\Re \int_{-\pi}^{\pi} \left(\frac{1}{J\sqrt{n}} \sum_{j=1}^J \sum_{|l| \leq n} f_{l-\theta_j^*} e^{-i\lambda(l-\theta_j)} \right) \left(\frac{1}{J\sqrt{n}} \sum_{j=1}^J \sum_{|l| \leq n} \sigma_l \epsilon_{j,l} e^{-i\lambda(l-\theta_j)} \right) \frac{d\lambda}{2\pi}$$

- $Q_n(\theta) = \int_{-\pi}^{\pi} \left| \frac{1}{J\sqrt{n}} \sum_{j=1}^J \sum_{|l| \leq n} \sigma_l \epsilon_{j,l} e^{-i\lambda(l-\theta_j)} \right|^2 \frac{d\lambda}{2\pi}$ est une forme quadratique du bruit.

Comme l'autopériodogramme converge étroitement vers la distribution spectrale F , le premier terme de D_n (5.9) converge simplement vers M . On étudie maintenant la convergence uniforme de ce terme (5.9) vers M . Soit G_n la fonction définie comme il suit :

$$G_n(\theta) = \left| \int_{-\pi}^{\pi} |\phi(\lambda, \theta)|^2 I_n(\lambda) d\lambda - \int_{-\pi}^{\pi} |\phi(\lambda, \theta)|^2 dF(\lambda) \right|.$$

Puisque la fonction G_n est continue sur l'ensemble compact Θ , elle est bornée et atteint ses bornes. Il existe donc θ_n tel que :

$$G_n(\theta_n) = \sup_{\theta \in \Theta} G_n(\theta).$$

Démontrer la convergence uniforme de (5.9) revient à démontrer que $G_n(\theta_n)$ converge vers 0. Comme $(G_n(\theta_n))_n$ est une suite positive, il suffit de montrer que de toute suite extraite de $(G_n(\theta_n))$ on peut extraire une sous-suite convergente vers 0. Notons $(n_k)_{k \geq 1}$ une sous-suite de (n) , et $\tilde{\theta}$ sa limite. Comme Θ est un ensemble compact, on peut extraire de la suite $(\theta_{n_k})_k \subset \Theta$ une sous-suite convergente. Notons $(\theta_{n_{k(p)}})_p$ cette sous-suite. Alors on a :

$$G_{n_{k(p)}}(\theta_{n_{k(p)}}) = G_{n_{k(p)}}(\theta_{n_{k(p)}}) - G_{n_{k(p)}}(\tilde{\theta}) + G_{n_{k(p)}}(\tilde{\theta}).$$

Puisque (5.9) converge simplement vers M sur Θ , $G_{n_{k(p)}}(\tilde{\theta})$ converge vers 0. De plus, on a l'inégalité suivante :

$$\left| G_{n_{k(p)}}(\theta_{n_{k(p)}}) - G_{n_{k(p)}}(\tilde{\theta}) \right| \leq \int_{-\pi}^{\pi} \left| |\phi(\lambda, \theta_{n_{k(p)}})| - |\phi(\lambda, \tilde{\theta})|^2 \right| (I_n(\lambda) d\lambda + dF(\lambda)).$$

Or ϕ est uniformément continue sur $[-\pi, \pi] \times \Theta$. Puisque $(\theta_{n_{k(p)}})_p$ converge, on en déduit que $G_{n_{k(p)}}(\theta_{n_{k(p)}}) - G_{n_{k(p)}}(\tilde{\theta})$ converge vers 0. Donc (5.9) converge uniformément vers M .

Définissons les fonctions suivantes :

$$D_{n,1}(\theta) = \int_{-\pi}^{\pi} \Re \left\{ \phi(\lambda, \theta) \sum_l J_n(\lambda) \frac{1}{J} \sum_{j=1}^J e^{i\lambda(\theta_j - \theta_j^*)} R_n(\theta_j^*) \right\} \frac{d\lambda}{2\pi},$$

$$D_{n,2}(\theta) = \int_{-\pi}^{\pi} \left| \frac{1}{J} \sum_{j=1}^J e^{i\lambda(\theta_j - \theta_j^*)} R_n(\theta_j^*) \right|^2 \frac{d\lambda}{2\pi}.$$

Maintenant, nous montrons que $D_{n,1}$, $D_{n,2}$, Q_n et L_n convergent uniformément en probabilité vers 0.

D'après l'inégalité de Cauchy-Schwarz sur l'espace de Hilbert des suites de carré sommable $l^2(\mathbb{Z})$, on a tout d'abord que :

$$\sup_{\theta \in \Theta} |D_{n,2}(\theta)| \leq \frac{1}{J} \sum_{j=1}^J \int_{-\pi}^{\pi} |R_n(\theta_j^*)|^2 d\lambda,$$

Comme par définition $\sum_{|l| \leq n} f_l^2/n$ converge vers $\gamma(0) = \int_{-\pi}^{\pi} dF(\lambda)$, $D_{n,2}$ converge uniformément vers zéro.

Par ailleurs, par l'inégalité de Cauchy-Schwarz sur l'espace des fonctions de carré intégrable sur $[-\pi, \pi]$, $\mathbb{L}^2[-\pi, -\pi]$, et en utilisant le fait que ϕ est bornée par 1, on a que :

$$\sup_{\theta \in \Theta} |D_{n,1}(\theta)| \leq 2 \left(\sup_{\theta \in \Theta} |D_{n,2}(\theta)| \right)^{1/2} \left(\int_{-\pi}^{\pi} I_n(\lambda) d\lambda \right)^{1/2}.$$

Ainsi de l'uniforme convergence de $D_{n,2}$, on déduit l'uniforme convergence vers zéro de $D_{n,1}$.

D'après l'inégalité de Cauchy-Schwarz dans l'espace $l^2(\mathbb{Z})$ et par l'égalité de Parseval dans l'espace $\mathbb{L}^2[-\pi, \pi]$, la fonction Q_n est uniformément bornée par,

$$Q_n^B = \sum_{j=1}^J \frac{1}{n} \sum_{|l| \leq n} \sigma_l^2 \epsilon_{j,l}^2.$$

L'hypothèse (5.2) garantit que Q_n^B converge vers zéro en probabilité. Donc Q_n converge uniformément en probabilité vers zéro.

Pour la fonction L_n , l'inégalité de Cauchy-Schwarz dans $\mathbb{L}^2[-\pi, \pi]$ établit que :

$$|L_n(\theta)| \leq 2Q_n(\theta)^{1/2} D_n^{1/2}.$$

On conclut que L_n converge uniformément en probabilité vers zéro du fait que Q_n converge uniformément vers zéro et D_n est uniformément borné. \square

Lemme 5.2 (Unicité du maximum). *Sous les hypothèses du théorème 5.1, M possède un unique maximum en θ^* .*

Démonstration. D'après l'inégalité de Cauchy-Schwarz, la fonction $|\phi|^2$ (voir les notations du lemme 5.1) est bornée par 1. Par conséquent, la fonction M atteint son maxima si, et seulement si,

$$|\phi(\cdot, \theta)|^2 = 1 \quad F\text{-presque sûrement.}$$

D'après la contrainte d'identifiabilité supposée sur le modèle ($\theta_1 = 0$) et le cas d'égalité de Cauchy-Schwarz, il faut et il suffit que pour tout λ F -presque partout,

$$\forall j = 2, \dots, J \quad e^{i\lambda(\theta_j - \theta_j^*)} = 1.$$

Or l'intérieur du support de F est non vide, on en déduit nécessairement que $\theta = \theta^*$. \square

Lemmes pour le théorème 5.2

Lemme 5.3. *Sous les hypothèses 5.2, on a que*

$$n^{(1-\alpha)/2} \nabla M_n(\theta^*) = \mathcal{O}_{\mathbb{P}}(1).$$

Démonstration. On montre d'abord que $n^{(1-\alpha)/2} \nabla D_n(\theta^*)$ converge vers zéro. Calculons les dérivées partielles pour $k = 2, \dots, J$:

$$\begin{aligned} \frac{d}{d\theta_k} D_n(\theta^*) &= 2\Re \int_{-\pi}^{\pi} \frac{-i\lambda}{J} \left(J_n(\lambda) + \frac{1}{J} \sum_{j=1}^J R_n(\theta_j^*) \right) \left(\overline{J_n(\lambda)} + \overline{R_n(\theta_k^*)} \right) \frac{d\lambda}{2\pi} \\ &= \partial_k D_{n,1}(\theta^*) + \partial_k D_{n,2}(\theta^*), \\ \text{où } \partial_k D_{n,1}(\theta^*) &= \frac{2}{J} \Re \int_{-\pi}^{\pi} i\lambda J_n(\lambda) \left(\frac{1}{J} \sum_{j=1}^J \overline{R_n(\theta_j^*)} - \overline{R_n(\theta_k^*)} \right) \frac{d\lambda}{2\pi} \\ \partial_k D_{n,2}(\theta^*) &= -\frac{2}{J} \Re \int_{-\pi}^{\pi} i\lambda \overline{R_n(\theta_k^*)} \frac{1}{J} \sum_{j=1}^J R_n(\theta_j^*) \frac{d\lambda}{2\pi}. \end{aligned}$$

Par l'inégalité de Cauchy-Schwarz, le terme $n^{(1-\alpha)/2} \partial_k D_{n,1}(\theta^*)$ est borné par,

$$2\pi \left\{ n^{1-\alpha} \int_{-\pi}^{\pi} \left| \frac{1}{J} \sum_{j=1}^J R_n(\theta_j^*) - R_n(\theta_k^*) \right|^2 \frac{d\lambda}{2\pi} \right\}^{1/2} \left\{ \int_{-\pi}^{\pi} I_n(\lambda) \frac{d\lambda}{2\pi} \right\}^{1/2}.$$

Ainsi pour montrer que $n^{(1-\alpha)/2} \partial_k D_{n,1}(\theta^*)$ converge vers zéro, il suffit que pour $k = 2 \dots J$ le terme suivant converge vers zéro,

$$n^{1-\alpha} \int_{-\pi}^{\pi} |R_n(\theta_k^*)|^2. \quad (5.12)$$

Par exemple si θ_k^* est positif, on a

$$\begin{aligned} n^{1-\alpha} \int_{-\pi}^{\pi} |R_n(\theta_k^*)|^2 \frac{d\lambda}{2\pi} &= \frac{1}{n^\alpha} \sum_{l=n-\theta_k^*+1}^n |f_l|^2 + \frac{1}{n^\alpha} \sum_{l=-n-\theta_k^*}^{-n-1} |f_l|^2, \\ &\leq \frac{2|\theta_k^*|}{n^\alpha} \sup_{l \in \mathbb{Z}} |f_l|^2. \end{aligned}$$

Comme le signal $(f_l)_l$ est borné, on déduit que (5.12) converge vers zéro.

D'autre part, la convergence vers zéro de $n^{(1-\alpha)/2}\partial_k D_{n,2}(\theta^*)$ est équivalente à la convergence vers zéro des termes suivants :

$$n^{(1-\alpha)/2} \Re \int_{-\pi}^{\pi} i\lambda \overline{R_n(\theta_k^*)} R_n(\theta_j^*) \frac{d\lambda}{2\pi}, \quad \text{pour tout } j \neq k.$$

Or d'après l'inégalité de Cauchy-Schwarz, ce terme est borné par

$$2\pi \left\{ n^{(1-\alpha)/2} \int_{-\pi}^{\pi} |R_n(\theta_k^*)|^2 \frac{d\lambda}{2\pi} \right\}^{1/2} \left\{ n^{(1-\alpha)/2} \int_{-\pi}^{\pi} |R_n(\theta_j^*)|^2 \frac{d\lambda}{2\pi} \right\}^{1/2}.$$

Ainsi on déduit du comportement asymptotique de (5.12) que $n^{(1-\alpha)/2}\partial_k D_{n,1}(\theta^*)$ et $n^{(1-\alpha)/2}\partial_k D_{n,2}(\theta^*)$ convergent vers zéro. Donc $n^{(1-\alpha)/2}\nabla D_n(\theta^*)$ converge vers zéro.

Étudions maintenant la convergence en probabilité vers zéro de $n^{(1-\alpha)/2}\nabla Q_n(\theta^*)$. Les dérivées partielles de Q_n sont pour $k = 2, \dots, J$:

$$\frac{d}{d\theta_k^*} Q_n(\theta^*) = \frac{2}{J^2 n} \sum_{j \neq k} \sum_{|l|, |p| \leq n} \epsilon_{j,l} \epsilon_{k,p} \sigma_l \sigma_p \Re \int_{-\pi}^{\pi} i\lambda e^{-i\lambda(p-l+\theta_j^*-\theta_k^*)} \frac{d\lambda}{2\pi}.$$

D'après l'inégalité de Cauchy-Schwarz, la variance est inférieure à :

$$\begin{aligned} \text{var} \left(n^{(1-\alpha)/2} \frac{d}{d\theta_k^*} Q_n(\theta^*) \right) &= \frac{4}{J^4 n^{1+\alpha}} \sum_{j \neq k} \sum_{|l|, |p| \leq n} \frac{\sigma_l^2 \sigma_p^2}{(p-l+\theta_j^*-\theta_k^*)^2}, \\ &\leq \frac{4}{J^4 n^{1+\alpha}} \sum_{j \neq k} \sum_{|l| \leq n} \sigma_l^2 \sum_{p \in \mathbb{Z}^*} \frac{1}{p^2}. \end{aligned}$$

L'hypothèse (5.2) implique donc la convergence vers zéro de la variance. Par conséquent, $n^{(1-\alpha)/2}\nabla Q_n(\theta^*)$ converge en probabilité vers zéro.

Il reste maintenant à étudier la comportement asymptotique de $n^{(1-\alpha)/2}\nabla L_n(\theta^*)$. Les dérivées partielles de L_n pour $k = 2, \dots, J$ sont :

$$\frac{d}{d\theta_k^*} L_n(\theta^*) = \frac{2}{J} \Re \int_{-\pi}^{\pi} \frac{i\lambda J_n(\lambda)}{\sqrt{n}} \left\{ \frac{1}{J} \sum_{j=1}^J \sum_{|l| \leq n} \sigma_l \epsilon_{j,l} e^{i\lambda(l-\theta_j^*)} - \sum_{|l| \leq n} \sigma_l \epsilon_{k,l} e^{i\lambda(l-\theta_k^*)} \right\} \frac{d\lambda}{2\pi} \quad (5.13)$$

$$+ \frac{2}{J} \Re \int_{-\pi}^{\pi} i\lambda R_n(\theta_k^*) \frac{1}{J\sqrt{n}} \sum_{j=1}^J \sum_{|l| \leq n} \sigma_l \epsilon_{j,l} e^{i\lambda(l-\theta_j^*)} \frac{d\lambda}{2\pi} \quad (5.14)$$

$$- \frac{2}{J} \Re \int_{-\pi}^{\pi} i\lambda \frac{1}{J} \sum_{j=1}^J R_n(\theta_j^*) \frac{1}{\sqrt{n}} \sum_{|l| \leq n} \sigma_l \epsilon_{k,l} e^{i\lambda(l-\theta_k^*)} \frac{d\lambda}{2\pi}. \quad (5.15)$$

On étudie en premier lieu le terme $n^{(1-\alpha)/2}$ (5.14). D'après l'inégalité de Cauchy-Schwarz, ce terme est borné par :

$$\begin{aligned} & \frac{2\pi}{J} \left(n^{1-\alpha} \int_{-\pi}^{\pi} |R_n(\theta_k^*)| \frac{d\lambda}{2\pi} \right)^{1/2} \sum_{j=1}^J \left(\int_{-\pi}^{\pi} \left| \sum_{|l| \leq n} \frac{\sigma_l \epsilon_{j,l} e^{i\lambda(l-\theta_j^*)}}{\sqrt{n}} \right|^2 \right)^{1/2} \\ & \leq \frac{2\pi}{J} \left(n^{1-\alpha} \int_{-\pi}^{\pi} |R_n(\theta_k^*)| \frac{d\lambda}{2\pi} \right)^{1/2} \sum_{j=1}^J \left(\sum_{|l| \leq n} \frac{\sigma_l^2 \epsilon_{j,l}^2}{n} \right)^{1/2}. \end{aligned}$$

On remarque que pour tout $j = 1 \dots J$, la variable suivante converge en probabilité vers zéro (à cause de l'hypothèse (5.2))

$$\frac{1}{n} \sum_{|l| \leq n} \sigma_l^2 \epsilon_{j,l}^2.$$

De plus, on a vu que (5.12) converge vers zéro. Donc $n^{(1-\alpha)/2}$ (5.14) converge en probabilité vers zéro. De même, la variable $n^{(1-\alpha)/2}$ (5.15) converge en probabilité vers zéro.

Étudions maintenant le terme $n^{(1-\alpha)/2}$ (5.13). On va montrer qu'il est uniformément tendue. Remarquons que (5.13) est une combinaison linéaire des variables $L_{n,j}$, $j = 1 \dots J$,

$$L_{n,j} = \Re \int_{-\pi}^{\pi} i\lambda J_n(\lambda) \frac{1}{\sqrt{n}} \sum_{|l| \leq n} \sigma_l \epsilon_{j,l} e^{i\lambda(l-\theta_j^*)} \frac{d\lambda}{2\pi}.$$

Pour un j fixé, L_n^j est une variable centrée. Il suffit alors de montrer que la variance de $n^{(1-\alpha)/2} L_n^j$ soit uniformément bornée :

$$\begin{aligned} \text{var}(n^{(1-\alpha)/2} L_n^j) &= \frac{1}{n^\alpha} \sum_{|l| \leq n} \sigma_l^2 \left(\Re \int_{-\pi}^{\pi} i\lambda J_n(\lambda) e^{i\lambda(l-\theta_j^*)} \frac{d\lambda}{2\pi} \right)^2 \\ &\leq \frac{\pi}{n^\alpha} \sum_{|l| \leq n} \sigma_l^2 \int_{-\pi}^{\pi} I_n(\lambda) \frac{d\lambda}{2\pi}. \end{aligned}$$

De la convergence étroitement de I_n vers F , et de l'hypothèse (5.8), on déduit que la variance est bien bornée. \square

Lemme 5.4. *Sous les hypothèses du théorème 5.2, la matrice hessienne converge uniformément en probabilité sur l'ensemble $\{\theta \in \Theta, \|\theta - \theta^*\| \leq \|\hat{\theta}_n - \theta^*\|\}$ vers la matrice carrée inversible $\nabla^2 M(\theta^*)$.*

Démonstration. Comme dans le lemme 5.1, $\nabla^2 M_n$ converge uniformément en probabilité sur Θ vers $\nabla^2 M$. On conclut, de la même façon que pour la proposition 2.6, que

$\nabla^2 M_n$ converge uniformément en probabilité sur $\{\theta \in \Theta, \|\theta - \theta^*\| \leq \|\hat{\theta}_n - \theta^*\|\}$ vers $\nabla^2 M(\theta^*)$. Il reste à vérifier que la matrice $\nabla^2 M(\theta^*)$ est inversible. Après calculs, cette matrice vaut,

$$\nabla^2 M(\theta^*) = \frac{2}{J} \int_{-\pi}^{\pi} \lambda^2 dF(\lambda) (I_{J-1} - \frac{J-1}{J} \mathbb{I}_{J-1} \mathbb{I}_{J-1}^T),$$

où I_{J-1} est la matrice identité de taille $J-1$, et $\mathbb{I}_{J-1} = (1, \dots, 1)^T$ est un vecteur de \mathbb{R}^{J-1} . Or si l'intérieur du support de F est non vide, le réel $\int_{-\pi}^{\pi} \lambda^2 dF(\lambda)$ est non nul. D'autre part, la matrice $(I_{J-1} - \frac{J-1}{J} \mathbb{I}_{J-1} \mathbb{I}_{J-1}^T)$ est inversible :

$$(I_{J-1} - \frac{J-1}{J} \mathbb{I}_{J-1} \mathbb{I}_{J-1}^T)^{-1} = (I_{J-1} + \frac{J-1}{J^2 - J + 1} \mathbb{I}_{J-1} \mathbb{I}_{J-1}^T).$$

□

Bibliographie

- A. Averbuch, R. R. Coifman, D. L. Donoho, M. Elad et M. Israeli. Fast and accurate polar Fourier transform. *Appl. Comput. Harmon. Anal.*, 21:145–167, 2006.
- S. Bayomog, X. Guyon, C. Hardouin et J. Yao. Test de différence de contrastes et somme pondérée de khi-deux. *Canad. J. Statist.*, 24(1):115–130, 1996. ISSN 0319-5724.
- R. Beran. Prepivoting test statistics : a bootstrap view of asymptotic refinements. *J. Amer. Statist. Assoc.*, 83(403):687–697, 1988. ISSN 0162-1459.
- B. Bercu et C. Prieur. Spectral properties of chaotic processes. *Stoch. Dyn.*, 6(3):355–371, 2006. ISSN 0219-4937.
- P. J. Bickel, A. J. Klaassen, Y. Ritov et J. A. Wellner. Efficient and adaptive inference in semi-parametric models, 1993.
- P. Bloomfield. *Fourier analysis of time series*. Wiley Series in Probability and Statistics : Applied Probability and Statistics. Wiley-Interscience [John Wiley & Sons], New York, second édition, 2000. ISBN 0-471-88948-2. An introduction.
- D. R. Brillinger. *Time series*, volume 36 de *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001. ISBN 0-89871-501-6. Data analysis and theory, Reprint of the 1981 edition.
- L. D. Brown et M. G. Low. Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.*, 24(6):2384–2398, 1996. ISSN 0090-5364.
- L. G. Brown. A survey of image registration techniques. *ACM Comput. Surv.*, 24(4):325–376, 1992. ISSN 0360-0300.
- E. J. Candès et D. L. Donoho. Recovering edges in ill-posed inverse problems : optimality of curvelet frames. *Ann. Statist.*, 30(3):784–842, 2002. ISSN 0090-5364. Dedicated to the memory of Lucien Le Cam.

- I. Castillo. Semiparametric second order efficient estimation of the period of a signal and applications. *preprint*, 2005.
- J. Coursol. *Technique statistique des modèles linéaires*. CIMPA, 1980.
- D. Dacunha-Castelle et M. Duflo. *Probability and statistics. Vol. II*. Springer-Verlag, New York, 1986. ISBN 0-387-96213-1. Translated from the French by David McHale.
- A. Dalalyan. Stein Shrinkage with Penalization and Second Order Efficiency in Semiparametrics. *Arxiv preprint math.ST/0509230*, 2005.
- A. S. Dalalyan, G. K. Golubev et A. B. Tsybakov. Penalized Maximum Likelihood and Semiparametric Second Order Efficiency. *Ann. Statist.*, 34(1):169–201, 2006.
- E. De Castro et C. Morandi. Registration of translated and rotated images using finite Fourier transforms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(5):700–703, 1987.
- S. Derrode et F. Ghorbel. Shape analysis and symmetry detection in gray-level objects using the analytical Fourier-Mellin representation. *Signal Processing*, 84(1):25–39, 2004.
- D. L. Donoho et I. M. Johnstone. Asymptotic minimaxity of wavelet estimators with sampled data. *Statist. Sinica*, 9(1):1–32, 1999. ISSN 1017-0405.
- I. L. Dryden et K. V. Mardia. *Statistical shape analysis*. Wiley Series in Probability and Statistics : Probability and Statistics. John Wiley & Sons Ltd., Chichester, 1998. ISBN 0-471-95816-6.
- B. Efron. Bootstrap methods : another look at the jackknife. *Ann. Statist.*, 7(1):1–26, 1979. ISSN 0090-5364.
- B. Efron et R. J. Tibshirani. *An introduction to the Bootstrap*. Chapman and Hall, 1993.
- N. I. Fisher et P. Hall. On bootstrap hypothesis testing. *Austral. J. Statist.*, 32(2):177–190, 1990. ISSN 0004-9581.
- F. Gamboa, J. M. Loubes et E. Maza. Shifts estimation for high dimensional data. Technical report of Laboratoire de statistique et probabilités, 2005.
- E. Gassiat et C. Lévy-Leduc. Efficient Semiparametric Estimation of the Periods in a Superposition of Periodic Functions with Unknown Shape. *Journal of Time Series Analysis*, 27(6):877–910, 2006.
- J. Gauthier, G. Bornard et M. Silbermann. Motions and pattern analysis : harmonic analysis on motion groups and their homogeneous spaces. *IEEE Trans. Systems Man Cybernet.*, 21(1):159–172, 1991. ISSN 0018-9472.

- F. Ghorbel. A complete invariant description for gray-level images by the harmonic analysis approach. *Pattern Recognition Letters*, 15(10):1043–1051, 1994.
- C. A. Glasbey. A review of image-warping methods. *Journal of Applied Statistics*, 25(2): 155–171, 1998.
- C. A. Glasbey et K. V. Mardia. A penalized likelihood approach to image warping. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 63(3):465–514, 2001. ISSN 1369-7412.
- V. Guardabasso, D. Rodbard et P. J. Munson. A versatile method for simultaneous analysis of families of curves. *The FASEB Journal*, 2:209–215, 1988.
- S. Haber. Numerical evaluation of multiple integrals. *SIAM Rev.*, 12:481–526, 1970. ISSN 1095-7200.
- J.V. Hajnal, D.L.G. Hill et D.J. Hawkes. *Medical Image Registration*. CRC Press, 2001.
- P. Hall. *The bootstrap and Edgeworth expansion*. Springer Series in Statistics. Springer-Verlag, New York, 1992. ISBN 0-387-97720-1.
- M. Hallin et B. J. M. Werker. Semi-parametric efficiency, distribution-freeness and invariance. *Bernoulli*, 9(1):137–165, 2003.
- P.R. Halmos. *A Hilbert Space Problem Book*. Springer, 1982.
- J. Horowitz. *Bootstrap Methods in Econometrics : Theory and Numerical Performance*. University of Iowa, College of Business Administration, Dept. of Economics, 1995.
- W. Härdle et J. S. Marron. Semiparametric comparison of regression curves. *The Annals of Statistics*, 18(1):63–59, 1990.
- R. Jane, H. Rix, P. Caminal et P. Laguna. Alignment methods for averaging of high-resolution cardiac signals : a comparative study of performance. *Biomedical Engineering, IEEE Transactions on*, 38(6):571–579, 1991.
- Y. Keller, A. Averbuch et M. Israeli. Pseudopolar-based estimation of large translations, rotations, and scalings in images. *IEEE Trans. Image Process.*, 14(1):12–22, 2005a. ISSN 1057-7149.
- Y. Keller, Y. Shkolnisky et A. Averbuch. The angular difference function and its application to image registration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(6):969–976, 2005b.
- A. Kneip et J. Engel. Model estimation in nonlinear regression under shape invariance. *Ann. Statist.*, 23(2):551–570, 1995. ISSN 0090-5364.

- A. Kneip et T. Gasser. Convergence and consistency results for self-modeling nonlinear regression. *The Annals of Statistics*, 16(1):82–112, 1988.
- A. Kneip et T. Gasser. Statistical tools to analyze data representing a sample of curves. *Ann. Statist.*, 20(3):1266–1305, 1992. ISSN 0090-5364.
- A. Kneip et T. Gasser. Searching for structure in curve samples. *J.Amer.Assoc.*, 90(432):1179–1188, 1995.
- A. Kneip, X. Li, K. B. MacGibbon et J. O. Ramsay. Curve registration by local regression. *Canad. J. Statist.*, 28(1):19–29, 2000. ISSN 0319-5724.
- C. D. Kuglin et D. C. Hines. The phase correlation image alignment method. *Proc. Int. Conf. on Cybernetics and Society*, pages 163–165, 1975.
- E. Laciár, R. Jane et D.H. Brooks. Improved alignment method for noisy high-resolution eeg and holter records using multiscale cross-correlation. *Biomedical Engineering, IEEE Transactions on*, 50(3):344–353, 2003.
- J C. Lagarias, J A. Reeds, M H. Wright et P E. Wright. Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM J. Optim.*, 9(1):112–147 (electronic), 1999. ISSN 1052-6234.
- W. M. Lawton, E. A. Sylvestre et M. G. Maggio. Self-modeling nonlinear regression. *Technometrics*, 14:513–532, 1972.
- R. Lenz. *Group theoretical methods in image processing*, volume 413 de *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, 1990. ISBN 3-540-52290-5.
- Artur O. Lopes et Sílvia R. C. Lopes. Convergence in distribution of the periodogram of chaotic processes. *Stoch. Dyn.*, 2(4):609–624, 2002. ISSN 0219-4937.
- J.M. Loubes, E. Maza, M. Lavielle et L. Rodríguez. Road trafficking description and short term travel time forecasting, with a classification method. *Canad. J. Statist.*, 34(3):475–491, 2006. ISSN 0319-5724.
- B. Markussen. A statistical approach to large deformation diffeomorphisms. *cvprw*, 12:181, 2004. ISSN 1063-6919.
- M. McGuire. An image registration technique for recovering rotation, scale and translation parameters. *NEC Research Institute Technical Report TR*, pages 98–018, 1998.
- B. McNeney et J. A. Wellner. Application of convolution theorems in semiparametric models with non-iid data. *J. Statist. Plann. Inference*, 91:441–480, 2000.

- J. Möcks, T. Pham Dinh et T. Gasser. Testing for homogeneity of noisy signals evoked by repeated stimuli. *Ann. Statist.*, 12(1):193–209, 1984. ISSN 0090-5364.
- S. A. Murphy et A. W. Van der Vaart. On Profile Likelihood. *Journal of the American Statistical Association*, 95(450), 2000.
- J.A. Nelder et R. Mead. A simplex method for function minimization. *Computer Journal*, 7(4):308–313, 1965.
- M. S. Pinsker. Optimal Filtration of Square-Integrable Signals in Gaussian Noise. *PROB. INFO. TRANSMISSION*, 16(2):120–133, 1980.
- D. Ramakrishnan et R. J. Valenza. *Fourier analysis on number fields*, volume 186 de *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1999. ISBN 0-387-98436-4.
- J. O. Ramsay et X. Li. Curve registration. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 60(2): 351–363, 1998. ISSN 1369-7412.
- C. Rao. Some statistical methods for the comparison of growth curves. *Biometrics*, 14:1–17, 1958.
- B. S. Reddy et B. N. Chatterji. An FFT-based technique for translation, rotation, and scale-invariant image registration. *Image Processing, IEEE Transactions on*, 5(8):1266–1271, 1996.
- S. I. Resnick. *Extreme values, regular variation, and point processes*, volume 4 de *Applied Probability. A Series of the Applied Probability Trust*. Springer-Verlag, New York, 1987. ISBN 0-387-96481-9.
- J. A. Richards. *Remote Sensing Digital Image Analysis : An Introduction*. Springer, 2005.
- W. Rudin. *Fourier analysis on groups*. Wiley Classics Library. John Wiley & Sons Inc., New York, 1990. ISBN 0-471-52364-X. Reprint of the 1962 original, A Wiley-Interscience Publication.
- H. Sakoe et S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoustics, Speech and Signal Processing*, 26(1):43–49, 1978.
- J. Segman, J. Rubinstein et Y.Y. Zeevi. The canonical coordinates method for pattern deformation : theoretical and computational considerations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(12):1171–1183, 1992.
- B. W. Silverman. Incorporating parametric effects into functional principal components analysis. *J. Roy. Statist. Soc. Ser. B*, 57(4):673–689, 1995. ISSN 0035-9246.

- H. Strasser. Tangent vectors for models with independent but non-identically distributed observations. *Statistics and Decisions*, 7:127–152, 1989.
- A. B. Tsybakov. *Introduction à l'estimation non-paramétrique*, volume 41 de *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin, 2004. ISBN 3-540-40592-5.
- A. W. Van der Vaart. *Asymptotic Statistics*, volume 3 de *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998. ISBN 0-521-49603-9 ; 0-521-78450-6.
- A. W. Van der Vaart. Semiparametric statistics. Dans *Lectures on probability theory and statistics (Saint-Flour, 1999)*, volume 1781 de *Lecture Notes in Math.*, pages 331–457, Berlin, 2002. Springer.
- P. Wahlberg et S. Salomonsson. Methods for alignment of multi-class signal sets. *Signal Process.*, 83(5):983–1000, 2003. ISSN 0165-1684.
- K. Wang et T. Gasser. Alignment of curves by dynamic time warping. *Ann. Statist.*, 25(3): 1251–1276, 1997. ISSN 0090-5364.
- Y. Wang et M. M. Brown. A flexible model for human circadian rhythms. *Biometrics*, 52: 588–596, 1996.

Appendices

Annexe A

Notions et notations utilisées

A.1 Rappels de probabilité et statistiques

A.1.1 Variables gaussiennes complexes

Si X est un d -vecteur aléatoire gaussien à valeur dans \mathbb{R}^d , de moyenne μ_X et de matrice de covariance Σ_{XX} , on écrit

$$X \sim \mathcal{N}_d(\mu_X, \Sigma_{XX}).$$

Dans cette thèse, on considère souvent des variables aléatoires X à valeurs complexes. On reprend les notations de [Brillinger \(2001\)](#). Pour une telle variable X , on considérera le vecteur aléatoire à valeur dans \mathbb{R}^{2r} suivant :

$$\begin{bmatrix} \Re X \\ \Im X \end{bmatrix} \sim \mathcal{N}_{2r} \left(\begin{bmatrix} \Re \mu_X \\ \Im \mu_X \end{bmatrix}, \frac{1}{2} \begin{bmatrix} \Re \Sigma_{XX} & -\Im \Sigma_{XX} \\ \Im \Sigma_{XX} & \Re \Sigma_{XX} \end{bmatrix} \right),$$

pour $\mu_X \in \mathbb{C}^d$ et Σ_{XX} une $d \times d$ -matrice hermitienne positive. On note alors $X \sim \mathcal{N}_d^{\mathbb{C}}(\mu_X, \Sigma_{XX})$. X est un vecteur Gaussien complexe de moyenne μ_X et de matrice de covariance Σ_{XX} , et on a :

$$\begin{aligned} \mathbb{E}(X) &= \mu_X, \\ \mathbb{E} \left[(X - \mu_X) \overline{(X - \mu_X)}^T \right] &= \Sigma_{XX}. \end{aligned}$$

A.1.2 Convergence Stochastique

Pour cette partie, le lecteur peut se reporter à [Van der Vaart \(1998\)](#).

Symboles stochastiques o et \mathcal{O} . La notation $o_{\mathbb{P}}(1)$ est une abréviation pour signifier qu'une variable aléatoire qui converge en probabilité vers zéro. La notation $\mathcal{O}_{\mathbb{P}}(1)$ signifie que la variable aléatoire est bornée en probabilité ou uniformément tendue. Plus généralement, pour une suite de variables aléatoires R_n ,

$$\begin{aligned} X_n = o_{\mathbb{P}}(R_n) \quad \text{signifie} \quad X_n = R_n Y_n \quad \text{avec} \quad Y_n = o_{\mathbb{P}}(1); \\ X_n = \mathcal{O}_{\mathbb{P}}(R_n) \quad \text{signifie} \quad X_n = R_n Y_n \quad \text{avec} \quad Y_n = \mathcal{O}_{\mathbb{P}}(1). \end{aligned}$$

Lemme A.1. Soit R une fonction définie sur un ouvert de \mathbb{R}^d telle que $R(0) = 0$. Soit X_n une suite de variables aléatoires à valeur dans le domaine de R qui converge en probabilité vers 0. Alors pour tout $p > 0$,

- si $R(h) = o(\|h\|^p)$ quand $h \rightarrow 0$, alors $R(X_n) = o_{\mathbb{P}}(\|X_n\|^p)$;
- si $R(h) = \mathcal{O}(\|h\|^p)$ quand $h \rightarrow 0$, alors $R(X_n) = \mathcal{O}_{\mathbb{P}}(\|X_n\|^p)$.

Théorème de Lindeberg-Feller

Théorème A.1 (Théorème de Lindeberg-Feller). Soient $Y_{n,1}, \dots, Y_{n,k_n}$ des vecteurs aléatoires indépendants de carré intégrable tels que :

$$\begin{aligned} \sum_{i=1}^{k_n} \mathbb{E} \|Y_{n,i}\|^2 1_{\{\|Y_{n,i}\| > \epsilon\}} &\rightarrow 0, \text{ pour tout } \epsilon > 0 \\ \sum_{i=1}^{k_n} \text{cov}(Y_{n,i}) &\rightarrow \Sigma. \end{aligned}$$

Alors la suite aléatoire $\sum_{i=1}^{k_n} (Y_{n,i} - \mathbb{E}(Y_{n,i}))$ converge en loi vers une loi normale $\mathcal{N}(0, \Sigma)$.

A.2 Espace de Hilbert

Un exposé plus détaillé peut être trouvé dans [Halmos \(1982\)](#). Un espace vectoriel réelle \mathcal{H} est appelé espace préhilbertien quand il est muni d'une forme bilinéaire définie positive $\langle \cdot, \cdot \rangle_{\mathcal{H}} : \mathcal{H}^2 \rightarrow \mathbb{R}$, appelé produit scalaire. On note $\|\cdot\|_{\mathcal{H}} : h \in \mathcal{H} \rightarrow \|h\|_{\mathcal{H}} = \langle h, h \rangle_{\mathcal{H}}^{1/2} \in \mathbb{R}_+$ la norme associée.

Inégalité de Cauchy-Schwarz. Pour tout $h \in \mathcal{H}$ et $h' \in \mathcal{H}$,

$$|\langle h, h' \rangle_{\mathcal{H}}| \leq \|h\|_{\mathcal{H}} \|h'\|_{\mathcal{H}}.$$

L'inégalité précédente est une égalité si, et seulement si h et h' sont colinéaires, c'est à dire qu'il existe une constante $\lambda \in \mathbb{R}$ telle que $h = \lambda h'$.

Orthogonalité. Lorsque pour $h \in \mathcal{H}$ et $h' \in \mathcal{H}$, on a $\langle h, h' \rangle_{\mathcal{H}} = 0$, on dit que h est orthogonal à h' , et l'on écrit parfois $h \perp h'$. Supposons que \mathcal{H} soit complet (on dit alors que \mathcal{H} est un espace de Hilbert). Si E est un sous-espace de \mathcal{H} , on note E^\perp le sous-espace fermé des éléments $h \in \mathcal{H}$ qui sont orthogonaux à tous les $e \in E$.

De plus si E est un sous-espace fermé de \mathcal{H} , on a :

- Tout $h \in \mathcal{H}$ se décompose d'une façon unique :

$$h = \Pi_E h + (1 - \Pi_E)h$$

en une somme de $\Pi_E h \in E$ et $(1 - \Pi_E)h \in E^\perp = \{h \in \mathcal{H}, \forall e \in E, h \perp e\}$.

- L'application Π_E est linéaire, elle appelée projection orthogonale de \mathcal{H} sur E , et on a

$$\|h\|_{\mathcal{H}}^2 = \|\Pi_E h\|_{\mathcal{H}}^2 + \|(1 - \Pi_E)h\|_{\mathcal{H}}^2.$$

L'application $h \rightarrow \langle h, h' \rangle_{\mathcal{H}}$ est, pour tout $h' \in \mathcal{H}$ une forme linéaire continue sur \mathcal{H} . En fait toutes les formes linéaire continue sur \mathcal{H} sont de ce type.

Théorème A.2 (Théorème de représentation de Riesz). *Si b^* est une forme continue linéaire sur \mathcal{H} , il existe un unique $h_{b^*} \in \mathcal{H}$ tel que*

$$\forall h \in \mathcal{H}, \quad b^*(h) = \langle h_{b^*}, h \rangle_{\mathcal{H}}.$$

Matrice de Gramm. Soit \mathcal{H} un espace préhilbertien réel, et h_1, \dots, h_n n vecteurs de \mathcal{H} . On appelle *matrice de Gramm* de h_1, \dots, h_n la matrice carrée de taille n $(\langle h_i, h_j \rangle_{\mathcal{H}})_{1 \leq i, j \leq n}$. On note cette matrice $\text{Gramm}(h_1, \dots, h_n)$.

Proposition A.1. *Toute matrice de Gramm est symétrique positive. De plus la matrice de Gramm de n vecteurs h_1, \dots, h_n est inversible si, et seulement si la famille $\{h_1, \dots, h_n\}$ est libre, c'est à dire qu'aucun vecteur de cette famille n'est combinaison linéaire des autres.*

A.3 Produit tensoriel de deux espaces de dimensions finies

Ce rappel a été rédigé à partir de l'ouvrage de Coursol (1980).

Soient E et F deux \mathbb{R} -espaces vectoriels de dimensions respectives p et q . Soient $\{e_i\}_{i=1\dots p}$ une base de E , et $\{f_i\}_{i=1\dots q}$ une base de F . On appelle produit tensoriel des espaces E et F tout \mathbb{R} -espace vectoriel G de dimension pq de sorte qu'il existe une

application bilinéaire,

$$\begin{aligned}\phi : E \times F &\rightarrow G \\ (e_i, f_j) &\rightarrow g_{i,j}, \quad \forall i = 1 \dots p, \forall j = 1 \dots q,\end{aligned}$$

où $\{g_{i,j}\}_{i=1\dots p, j=1\dots q}$ est une base de G . On note alors $G = E \otimes F$. Et pour tout $x \in E$ et pour tout $y \in F$, on note $x \otimes y$ l'élément $g = \phi(x, y)$.

Si $E = \mathbb{R}^p$ et $F = \mathbb{R}^q$, on note encore $E \otimes F = \mathbb{R}^{p \otimes q}$. Cet espace est alors isomorphe à l'espace vectoriel des matrices réelles de dimension $p \times q$.

Soient a un endomorphisme de E , et $A = (a_{ij})$ sa matrice associée dans la base $\{e_i\}_{i=1\dots p}$. Soient b un endomorphisme de F , et B sa matrice associée dans la base $\{e_j\}_{j=1\dots q}$. Alors l'application, notée $a \otimes b$,

$$\begin{aligned}a \otimes b : \quad E \otimes F &\rightarrow E \otimes F \\ x \otimes y &\rightarrow a(x) \otimes b(y)\end{aligned}$$

induit un endomorphisme de $E \otimes F$. Donnons une représentation matricielle de cet endomorphisme. Ordonnons la base $\{e_i \otimes f_j\}_{i=1\dots p, j=1\dots q}$ de la façon suivante :

$$e_1 \otimes f_1, \dots, e_1 \otimes f_q, e_2 \otimes f_1, \dots, e_p \otimes f_1, \dots, e_p \otimes f_q.$$

La matrice associée à $a \otimes b$ dans la base $\{e_i \otimes f_j\}_{i=1\dots p, j=1\dots q}$ peut se définir par bloc de façon suivante :

$$A \otimes B = \begin{pmatrix} a_{1,1}B & a_{1,2}B & \dots & a_{1,p}B \\ a_{2,1}B & a_{2,2}B & \dots & a_{2,p}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{q,1}B & a_{q,2}B & \dots & a_{q,p}B \end{pmatrix}.$$

Proposition A.2. *Soient a un endomorphisme de l'espace vectoriel E , et b un endomorphisme de l'espace vectoriel F . L'endomorphisme $a \otimes b$ de $E \otimes F$ est un isomorphisme si, et seulement si a est un isomorphisme de E et b est un isomorphisme de F .*

Corollaire A.1. *Soient A une matrice carrée de taille p , et B une matrice carrée de taille q . La matrice $A \otimes B$ est inversible si, et seulement si les matrices A et B le sont.*

Annexe B

Programmes MATLAB

Sommaire

A.1	Rappels de probabilité et statistiques	187
A.2	Espace de Hilbert	188
A.3	Produit tensoriel de deux espaces de dimensions finies	189

Pour chacun des algorithmes, nous utilisons la fonction `fminsearch` de MATLAB. Cet algorithme calcule le minimum d'une fonction f . Il emploie la méthode de [Lagarias et al. \(1999\)](#).

B.1 Programmes relatifs au chapitre 2

La fonction `fo` est une fonction définie par l'utilisateur afin de générer un jeu de données.

B.1.1 Calcul des estimateurs

```
clear
% Déclaration des variables globales
global J n m L Y D
% Paramètres fixés par l'utilisateur
J=3; % Nombre de courbes
n=125; % Nombre d'observations par courbes
% Points d'observation pas=2*pi/n;
T=0 :pas : (2*pi-pas); % temps d'observation
```

```

% Choix du nombre de coefficients de Fourier
m=10;
L=[-m :1 :m];

% Génération des données
 $\theta^* = [0 \ 0.3 \ 0.8]$ ; % Le paramètre de translation
 $a^* = [0.75 \ 1.1990 \ 1]$ ; % Le paramètre d'échelle
 $v^* = [0 \ 1 \ 3]$ ; % Le paramètre de niveau
y=diag( $a^*$ )*fo(ones(J,1)*T- $\theta^{*'}*$ ones(1,n))+ $v^{*'}*$ ones(1,n); % Génération des données non bruitées
 $\sigma = 1$ ; % Le niveau de bruit
Y=y+ $\sigma*$ randn(J,n); % Addition d'un bruit blanc

% Calcul des coefficients de Fourier
D=(Y*exp(-i.*(T'*L)))./n;

% Calcul des estimateurs de  $\theta^*$ ,  $a^*$  et  $v^*$ 
 $\theta_0$ =zeros(1,J-1);
 $a_0$ =ones(1,J-1); % on considère ici le jeu de contraintes d'identifiabilité n°3
 $v_0$ =zeros(1,J-1);
 $\hat{\alpha}_n$ =fminsearch(@Ln,[ $\theta_0 \ a_0 \ v_0$ ],optimset('MaxFunEvals',200000,'MaxIter',1000000));
 $\hat{\theta}_n$ =[0  $\hat{\alpha}_n(1:(J-1))$ ]; % Estimateur des translations
 $\hat{a}_n$ =[sqrt(J-sum( $\hat{\alpha}_n(J:2*(J-1)).^2$ ))  $\hat{\alpha}_n(J:2*(J-1))$ ]; % Estimateur des échelles
 $\hat{v}_n$ =[ $\hat{\alpha}_n(2*J-1:3*J-2)$ ]; % Estimateur des niveaux

```

B.1.2 Définition du processus de contraste (vraisemblance profilée)

```

function M=Ln( $\alpha$ )
global L D J n Y
% Contraintes d'identifiabilité des paramètres
 $\theta$ =[0  $\alpha(1:J-1)$ ];
a=[0  $\alpha(J:(2*J-2))$ ];
a(1)=sqrt(J-sum(a.^2));
v=[ $\alpha(2*J-1:3*J-2)$ ];
% Calcul des coefficients de Fourier (2.25)

```

```

e=exp(i.*theta'*L);
c-hat=sum(diag(a)*e.*D)./J;
% Calcul du critère
P=(abs(c-hat).^2);
M=sum(sum((Y-v'*ones(1,n)).^2))./(n*J)-sum(P);

```

B.2 Programmes relatifs au chapitre 3

L'algorithme bootstrap utilise "*The Bootstrap Toolbox*" développée par Abdelhak M. Zoubir et D. Robert Iskander. Le processus de contraste est défini à partir de deux fonctions M_n (prenant en variable globale la transformée de Fourier des données D), et M_n^* (prenant en variable globale la transformée de Fourier des échantillons bootstrap D^*)

B.2.1 Algorithme bootstrap du test

```

clear;
global B J n T m L delta D D*
% Description des variables globales
B=200; % Nombre d'échantillons bootstrap générés
J=2; % Nombre de courbes
n=201; % Nombre d'observations par courbes
pas=2*pi/n;
T=0 :pas : (2*pi-pas); % Temps d'observation
m=10; % Nombre de coefficients de Fourier
L=[-m :1 :-1 1 :1 :m]; % Vecteurs des indices des coefficients de Fourier considérés
delta=ones(1,2*m); % Suite de poids pour les coefficients de Fourier, ici la troncature

% Génération des données
sigma=0.8; % niveau de bruit
theta=[0 0.8];
a=[0.75 1.1990];
nuv=[-1 0.5];
y=diag(a)*fo(ones(J,1)*T-theta'*ones(1,n))+nuv'*ones(1,n);
Y=y+sigma.*randn(J,n); % les données

```

% Calcul des coefficients de Fourier

$D = (Y * \exp(-i * (T' * L))) ./ n$; *% calcul des coefficients de Fourier pour $l \in$*

$D_0 = \text{sum}(Y')$; *% calcul des coefficients de Fourier pour $l = 0$*

% Estimation des résidus

$\hat{\epsilon} = \text{real}(Y - D * \exp(i * L' * T) - D_0 * \text{ones}(1, n))$; $\tilde{\epsilon} = \hat{\epsilon} - \text{sum}(\hat{\epsilon}') / (n * J)$; *% cen-*
trage des résidus

$\hat{\sigma}^2 = \text{sum}(\text{sum}((\tilde{\epsilon} . \wedge 2))) / (n * J)$; *% estimation de la variance*

% Estimation des paramètres

$\theta_0 = \text{zeros}(1, J-1)$;

$a_0 = \text{ones}(1, J-1)$; *% on considère ici le jeu de contraintes d'identifiabilité n°3*

$v_0 = \text{zeroszeros}(1, J-1)$;

$[\hat{\alpha}_n, \text{val}] = \text{fminsearch}(@\text{Mn}, [\theta_0 \ a_0 \ v_0], \text{optimset}('MaxFunEvals', 90000, 'MaxIter', 6000))$;

$\hat{\theta}_n = [0 \ \hat{\alpha}_n(1 : (J-1))]$; *% Estimateur des translations*

$\hat{\alpha}_n = [\text{sqrt}(J - \text{sum}(\hat{\alpha}_n(J : 2 * (J-1)) . \wedge 2)) \ \hat{\alpha}_n(J : 2 * (J-1))]$; *% Estimateur des échelles*

$\hat{v}_n = D_0$; *% Estimateur des niveaux*

% Calcul de la statistique de test

$T_n = n * (\text{sum}(\text{sum}(\text{abs}(D . \wedge 2))) ./ J + \text{val}) / \hat{\sigma}^2$;

% Génération des échantillons bootstrap

% Échantillon bootstrap des résidus

$\tilde{\epsilon}_1^* = \text{bootrsp}(\tilde{\epsilon}(1, :), 2 * B)$;

$\tilde{\epsilon}_2^* = \text{bootrsp}(\tilde{\epsilon}(2, :), 2 * B)$;

% Estimation des coefficients de Fourier de la forme commune

$\hat{c} = \text{sum}(\text{diag}(\hat{a}) * (\exp(i * \hat{\theta}' * L)) * D) ./ J$;

$y^* = \text{zeros}(J, n)$;

$y^*(1, :) = \text{real}(\hat{a}(1) * \hat{c} * \exp(i * L' * (T - \hat{\theta}(1) * \text{ones}(1, n)))) + \hat{v}(1)$; $y^*(2, :) = \text{real}(\hat{a}(2) * \hat{a} * \exp(i * L' * (T - \hat{\theta}(2) * \text{ones}(1, n)))) + \hat{v}(2)$;

% Échantillon bootstrap de la statistique de test

$T^* = \text{zeros}(1, B)$;

for b=1 :B;

$Y^*(1, :) = y^*(1, :) + \tilde{\epsilon}_1^*(:, b)'$;

$Y^*(2, :) = y^*(2, :) + \tilde{\epsilon}_2^*(:, b)'$;

$D^* = \text{zeros}(J, 2 * m)$;

```

D*=(Y*exp(-i.*(T'*L)))./n;
D0*=sum(Y*')./n; % Estimation bootstrap de la variance
ê*=real(Y*-D*exp(i.*L'*T)-D0'*ones(1,n));
ê*=ê*-sum(sum(ê*'))/(n*J).*ones(J,n);
ô2,*=sum(sum((ê*.2')))/(n*J);
% Calcul de la statistique de test bootstrap
θ0=zeros(1,J-1);
a0=ones(1,J-1); % on considère ici le jeu de contraintes d'identifiabilité n°3
v0=zeros(1,J-1);
[ân*, val*]=fminsearch(@Mn*,[θ0 a0 v0],optimset('MaxFunEvals',90000,'MaxIter',6000));
T*(1,b)=n*(sum(sum(abs(D*.2)))./J+val*)/ô2,*;
end;

% Estimateur bootstrap de la quantile d'ordre p=0.95
tri=sort(T*); % On tri l'échantillon,
Tp=tri(1,0.95*B); % Estimateur bootstrap de la quantile d'ordre p,
if Tp<Tobs,
display('on rejette H0 au niveau 1 - p');

```

B.2.2 Définition du processus de contraste

```

function M=Mn(α)
global L D J m
θ=[0 α(1:J-1)];
a=[0 α(J:(2*J-2))];
a(1)=sqrt(J-sum(a.2));
% Calcul des coefficients de Fourier (2.25)
ĉ=sum(diag(a)*exp(i.*θ'*L).*D)./J;
M=-sum(abs(ĉ).2);

```

B.3 Programmes relatifs au chapitre 5

```

% Déclaration des variables globales
global n J Y;
J=3; % Nombre de signaux
n=35; % Taille de l'échantillon par signaux : 2n+1
% Y correspond aux données, c'est une matrice J×(2n+1). Ces données sont rentrées

```

par l'utilisateur, ou générées par celui-ci

% Pour estimer les décalages, on peut soit avoir recourt à une visualisation graphique du contraste M ou soit utiliser la procédure `fminsearch`. $J \times (2n+1)$.

B.3.1 Définition du processus de contraste

Le processus de contraste est défini à partir d'une fonction `matlab`, nommée ici `M`. Cette fonction fait appelle à une autre fonction `Dsignal`. `Dsignal` calcule les transformées de Fourier rephasées des signaux. `M` utilise l'approximation de Riemann pour intégrer le module au carré de la moyenne synchronisée des transformées de Fourier.

Fonction `M`

```
function[y]=M(theta)
% Le résultat est un réel
global n J;
K=101; % Discrétisation de l'approximation de Riemann
z=0;
pi=3.14;
for t=-K :K
z=z+abs(mean(Dsignal(pi*t/K,theta))).^2;
end;
y=-z./K;
```

Fonction `Dsignal`

```
function[y]=Dsignal(lambda,theta) % Le résultat est un vecteur de taille J
global n J Y;
L=[1 :1 :n];
e=exp(-i*lambda.*L);
y=sum((Y.*(ones(J,1)*e))')./sqrt(n).*exp(i*lambda*[0 theta]);
```

Liste des symboles et abréviations

Abréviation	Description	Définition
$[x]$	Partie entière du réelle $x \in \mathbb{R}$	
$ c $	Module de $c \in \mathbb{C}$	
\bar{c}	Complexe conjugué de $c \in \mathbb{C}$	
$\langle \cdot, \cdot \rangle_{\mathcal{H}}$	Produit scalaire associé à l'espace de Hilbert \mathcal{H}	page 188
\otimes	Produit tensoriel de matrices ou de vecteurs	page 189
\cdot^T	Opérateur de transposition des matrices	
∇	Opérateur gradient	
∇^2	Opérateur hessien	
$\mathbb{1}_J$	Vecteur $(1, 1, \dots, 1)^T$ de \mathbb{R}^J	
AFMT	Analytical Fourier Mellin Transform	page 126
$\arg(c)$	Argument de $c \in \mathbb{C}$	
\mathbb{C}	Corps des nombres complexes	
Δ_n	Processus centré pour le développement LAN	page 16
DFT	Discrete Fourier Transform	page 33
diag	Matrice diagonale	
\mathbb{E}	Espérance mathématique	
\mathcal{G}	Espace des coordonnées polaires	
Gramm	Matrice de Gramm	page 189
i.i.d.	Indépendant identiquement distribué	
$\Im(c)$	Partie imaginaire de $c \in \mathbb{C}$	
$\mathbb{L}^p(\mathcal{D})$	Espace des fonction p -intégrables sur l'espace \mathcal{D} ($p \in \mathbb{N}^*$)	
$l^p(\mathbb{Z})$	Espace des séries p -sommables ($p \in \mathbb{N}^*$)	
LAN	Local Asymptotic Normality	page 15
\mathbb{N}	Ensemble des entiers naturels	
$\mathcal{N}^{\mathbb{C}}(0, 1)$	Variable gaussienne complexe	page 187

Abréviation	Description	Définition
$\mathcal{O}_{\mathbb{P}}$	Symboles stochastiques	page 188
$o_{\mathbb{P}}$	Symboles stochastiques	page 188
\mathbb{P}	Mesure de probabilité	
\mathbb{R}	Corps des nombres réels	
$\Re(c)$	Partie réelle de $c \in \mathbb{C}$	
$\mathcal{R.S.T.}$	Rotation, Scale and Translation	
SEMOR	Self-Modeling nonlinear Regression	page 26
\mathbb{T}	Le tore $\mathbb{R}/(2\pi\mathbb{Z})$	
var	Variance	
\mathbb{Z}	Anneau des entiers	

Table des figures

2.1	Illustration du modèle (2.3) avec $d = 1$ et $J = 2$ courbes. Le graphique (a) représente deux fonctions vérifiant (2.2). La figure (b) représente ces mêmes fonctions bruitées avec $\sigma^* = 0.8$. La forme commune est générée par la fonction $f^*(x) = 20(1 - x/(2\pi))x/(2\pi)$	30
2.2	Diagrammes de Tukey (a) pour le jeu n°3 de contraintes d'identifiabilité ($c_0(f^*)=0$), (b) pour le jeu n°2 de contraintes d'identifiabilité ($v_1^*=0$) . .	50
2.3	Modèle de translation, $J = 3$. Graphique de W_n	54
2.4	Modèle d'amplitude, $J = 3$: (a) graphique de W_n , (b) graphique de L_n . .	55
3.1	Les ondes P300 associées respectivement à l'effet placebo (a) et à l'effet du traitement (d). L'estimation des fonctions de régression dans le modèle avec $J = 22$ unités pour les ondes placebo (b) et les ondes traitements (e). L'estimation des fonctions de régression dans le modèle avec $J = 11$ courbes en considérant uniquement les ondes placebo (c) et les ondes traitement (f).	92
3.2	Puissance du test pour l'hypothèse nulle " $f_j^* = a_j^* f^*(\cdot - \theta_j^*) + v_j^*$ " avec $J = 2$ en fonction de la variable β . Chaque point est calculé à partir de 1000 procédures de test, et les paramètres sont $(\theta_1^*, \theta_2^*) = (0, 0.8)$, $(a_1^*, a_2^*) = (0.75, 1.19)$, $(v_1^*, v_2^*) = (-1, 0.5)$, $\sigma^* = 0.8$, et $n = 201$	97
3.3	(a) Puissance du test pour l'hypothèse nulle " $f_j^* = a_j^* f^*(\cdot - \theta_j^*)$ " avec $J = 3$ en fonction de v_2 . Chaque point est calculé à partir de 1000 procédures de test et les paramètres sont $(\theta_1^*, \theta_2^*, \theta_3^*) = (0, 0.27, 0.83)$, $(a_1^*, a_2^*, a_3^*) = (0.75, 1.19, 1)$, $\sigma^* = 0.8$, et $n = 201$. (b) Les données sont générées à partir de $Y_{i,1} = a_2^* f^*(t_i - \theta_2^*) + \sigma^* \epsilon_{i1}$, et $Y_{i,2} = a_2^* f^*(t_i - \theta_2^*) + 0.8 + \sigma^* \epsilon_{i2}$, $i = 1, \dots, n$	98

3.4	(a) Puissance du test pour l'hypothèse nulle $f_j^* = f^*(\cdot - \theta_j^*)$ avec $J = 3$ en fonction de a_2 . Chaque point est calculé à partir de 1000 procédures de test et les paramètres sont $(\theta_1^*, \theta_2^*, \theta_2^*) = (0, 0.27, 0.83)$, $\sigma^* = 0.8$, et $n = 201$. (b) Les données sont générées à partir de $Y_{i,1} = f(t_i - \theta_2^*) + \sigma^* \epsilon_{i1}$, et $Y_{i,2} = 1.3f(t_i - \theta_2^*) + \sigma^* \epsilon_{i2}$ $i = 1, \dots, n$	99
4.1	Les images (a),(b) et (c) représentent le même papillon observé avec des tailles, des orientations et des emplacements différents. Les images (d),(e) et (f) représentent un autre papillon de forme différente, avec des tailles, des orientations et des emplacements divers.	120
4.2	Images simulées f_j^* : (a) $j = 1$, (b) $j = 2$, (c) $j = 3$, (d) $j = 4$, (e) $j = 5$, (f) $j = 6$	136
4.3	Images bruitées avec $s2n = 1$: (a) $j = 1$, (b) $j = 2$, (c) $j = 3$, (d) $j = 4$, (e) $j = 5$, (f) $j = 6$	138
4.4	Alignement des images de papillons.	141
5.1	Graphique du processus de contraste (5.6) pour des variables indépendantes identiquement distribuées ($J = 3$ et $n = 62$) et pour différentes valeurs de σ : (a) $\sigma = 0$, (b) $\sigma = 0.2$, (c) $\sigma = 0.4$, (d) $\sigma = 0.6$	166
5.2	Diagrammes de Tukey pour les estimateurs de $\theta_3^* = 5$ (définis à partir du processus (5.6)) construits à partir de 100 jeux de données dans le modèle où les signaux sont indépendants et identiquement distribués	167
5.3	Graphique du processus de contraste (5.6) pour des signaux à moyenne mobile ($J = 3$ et $n = 62$) : (a) $MA(1)$ avec $(\beta_1, \beta_2, \beta_3) = (\sqrt{5}, 0, 0)$, (b) $MA(3)$ avec $(\beta_1, \beta_2, \beta_3) = (2, 0, -1)$	168
5.4	Graphique du processus de contraste (5.7) pour des signaux $AR(1)$ ($J = 3$ et $n = 62$) : (a) $\gamma = 0.3$ avec $w \equiv 1$, (b) $\gamma = 0.7$ avec $w \equiv 1$, $\gamma = 0.3$ avec $w(\lambda) = e^{-\lambda^2/4}$, $\gamma = 0.3$ avec $w(\lambda) = e^{-\lambda^2/2}$	169
5.5	Diagrammes de Tukey de l'estimateur de $\theta_3^* = 5$ (définis à partir du processus (5.7)) construits à partir de 100 jeux de données dans le modèle où les signaux sont des $AR(1)$ ($J = 3$ et $n = 62$) : (a) $\gamma = 0.3$ avec $w \equiv 1$, $\gamma = 0.3$ avec $w(\lambda) = e^{-\lambda^2/4}$, $\gamma = 0.3$ avec $w(\lambda) = e^{-\lambda^2/2}$	170

Liste des tableaux

2.1	Moyennes et écarts type (entre parenthèses) pour l'estimation des paramètres $\theta_2^* = 0.95$ et $a_2^* = 0.75$ calculés à partir de 1000 jeux de données. . .	56
3.1	Correspondance entre l'hypothèse nulle à tester et le processus de contraste associé.	97
3.2	Pourcentage de rejet de H_0 sous les hypothèses statistiques H_0 et H_1 , calculés à partir de 1000 tests avec $l = 1$	108
3.3	Pourcentages de rejet de H_0 sous les hypothèses H_0 et H_1 , calculés à partir de 1000 jeux de données et $m_n = 50$	110
4.1	Valeur des paramètres d'échelle, de rotation et de translation pour les $J = 6$ images	137
4.2	Moyennes empiriques et écarts-type (entre parenthèses) des estimateurs $\hat{a}_j, \hat{b}_j = (\hat{b}_j^1, \hat{b}_j^2), \hat{\theta}_j$ sur $M = 50$ jeux de données pour différent rapport signal-sur-bruit $s2n$. Les valeurs en gras représentent les valeurs à estimer des paramètres.	140

Résumé

Dans cette thèse, nous étudions des modèles semi-paramétriques dits de forme invariante. Ces modèles consistent en l'observation d'un nombre fixé de fonctions de régression identiques à un opérateur de déformation paramétriques près. Ce type de modèles trouve des applications dans les problèmes d'alignement de signaux continus (images 2D, rythmes biologiques, ...) ou discrets (electro-encéphalogramme, ...). Pour différents groupes de déformations, nous proposons des M -estimateurs pour les paramètres caractérisant les opérateurs associés aux fonctions de régression. Ces estimateurs minimisent ou maximisent des fonctions de contraste, construites à partir de la moyenne synchronisée des transformées de Fourier des données. De plus, pour l'un des modèles étudiés, nous prouvons l'efficacité semi-paramétrique de cet estimateur ainsi défini, et nous proposons un test d'adéquation du modèle de forme invariante construit à partir d'une des fonctions de contraste.

Abstract

The shape invariant model consist of the observation of a fixed number of regression functions which differ only by a parametric warping operator. This type of models finds applications in the problems of alignment of continuous signals (images 2D, circadian rhythms,...) or discrete (electroencephalogram,...). For various warping groups, we propose M -estimators for the parameters characterizing the warping operators associated with the regression functions. These estimators minimize or maximize criteria which are defined with the synchronized average of the Fourier transforms of the data. Moreover, for one of the studied models, we prove the semi-parametric efficiency of the proposed estimator, and we build a test of adequacy of the shape invariant model from one of the criteria.